

10/30/00

Please type a plus sign (+) inside this box → +

10-31-00 A

10  
C549 U  
10  
P  
P  
P

<b>UTILITY PATENT APPLICATION TRANSMITTAL</b>	
(Only for new nonprovisional applications under 37 CFR 1.53(b))	

Attorney Docket No. BB1168 US NA

First Named Inventor or Application Identifier

Edgar B. Cahoon et al.

Express Mail Label No. EK639605137US

Express Mailing Date October 30, 2000

10  
U.S. 69652  
10/30/00**APPLICATION ELEMENTS**

See MPEP chapter 600 concerning utility patent application contents.

1.  Fee (Authority to charge deposit account below.)  
(Submit an original, and a duplicate for fee processing)

2.  Specification [Total Pages 33]

- Descriptive title of the invention
- Cross References to Related Applications (if needed)
- Statement Regarding Fed sponsored R & D (if needed)
- Reference to Microfiche Appendix (if filed)
- Background of the Invention
- Brief Summary of the Invention
- Brief Description of the Drawings (if filed)
- Detailed Description
- Claim(s)
- Abstract of the Disclosure

3.  Drawing(s) (35 USC 113) [Total Sheets 0]

4.  Oath or Declaration [Total Pages 0]

- a.  Newly executed (original or copy)
- b.  Copy from a prior application (37 CFR 1.63(d))  
(for continuation/divisional with Box 14 completed)
- i.  **DELETION OF INVENTORS**  
Signed Statement below at 15 deleting  
inventor(s) named in the prior application,  
see 37 CFR 1.63(d)(2) and 1.33(b).

5.  Incorporation by Reference (*useable if Box 4b is checked*)  
The entire disclosure of the prior application, from which a  
copy of the oath declaration is supplied under Box 4b, is  
considered as being part of the disclosure of the  
accompanying application and is hereby incorporated by  
reference therein.

14. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:

Continuation    Divisional    Continuation-in-part (CIP)   of prior Application No.: PCT/US99/09280

15.  **DELETION OF INVENTOR(S) STATEMENT:** This application is being filed by less than all the inventors named in the prior application. In accordance with 37 CFR 1.63(d)(2) and 1.33(b), the Assistant Commissioner is requested to delete the name(s) of the following person or persons who are not inventors of the invention being claimed in this application:

16.  Amend the specification by inserting before the first line the sentence:

17.  Cancel in this application original claims \_\_\_\_\_ of the prior application before calculating the filing. (At least one original independent claim must be retained for filing purposes.)

18.  Priority of foreign Application No. \_\_\_\_\_ filed on \_\_\_\_\_ in

\_\_\_\_\_ is claimed under 35 U.S.C. 119.  
(country)

ADDRESS TO: Assistant Commissioner for Patents  
Box Patent Application  
Washington, DC 20231

- 6.  Microfiche Computer Program (Appendix)
- 7.  Nucleotide and/or Amino Acid Sequence Submission  
(if applicable, all necessary)
  - a.  Computer Readable Copy
  - b.  Paper Copy (identical to computer copy)  
Sequence Listing - 28 Pages
  - c.  Statement verifying identity of above copies

**ACCOMPANYING APPLICATION PARTS**

- 8.  Power of Attorney
- 9.  Information Disclosure Statement (IDS)/Cover Letter plus PTO-1449    Copies of IDS Citations
- 10.  Preliminary Amendment
- 11.  Return Receipt Postcard (MPEP 503)  
(Should be specifically itemized)
- 12.  Certified Copy of Priority Document(s)  
(if foreign priority is claimed)
- 13.  Other:  
-

CLAIMS	(1) FOR	(2) NUMBER FILED	(3) NUMBER EXTRA	(4) RATE	(5) CALCULATIONS
	<b>TOTAL CLAIMS</b> (37 CFR 1.16(e))	24 - 20 =	4	x \$ 18 =	<b>\$ 72.00</b>
	<b>INDEPENDENT CLAIMS</b> (37 CFR 1.16(b))	3 - 3 =	0	x \$ 80 =	<b>0</b>
	<b>MULTIPLE DEPENDENT CLAIM(S)</b> (if applicable)			+ \$ 270 =	<b>0</b>
				<b>BASIC FEE</b> (37 CFR 1.16(a))	+ <b>\$ 710.00</b>
				<b>TOTAL =</b>	<b>\$ 782.00</b>

19. The Commissioner is hereby authorized to credit overpayments or charge the following fees to Deposit Account No. 04-1928:

- Fees required under 37 CFR 1.16.
- Fees required under 37 CFR 1.17.

20.  Other:

**21. CORRESPONDENCE ADDRESS**

NAME	Kening Li			
ADDRESS	E. I. du Pont de Nemours and Company Legal - Patents 1007 Market Street			
CITY	Wilmington	STATE	Delaware	ZIP CODE
COUNTRY	U.S.A.	TELEPHONE	302-992-3749	FAX
				19898
				302-892-1026

**22. SIGNATURE OF ATTORNEY OR AGENT REQUIRED**

NAME	Kening Li	REG. NO.: 44,872
SIGNATURE		
DATE	30 October 2000	

EXPRESS MAIL LABEL NO: EK639605137US  
PATENT  
IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In the Application of:

EDGAR B CAHOON ET AL.

CASE NO.: BB1168 US NA

APPLICATION NO.: UNKNOWN

GROUP ART UNIT: UNKNOWN

FILED: CONCURRENTLY HEREWITH

EXAMINER: UNKNOWN

FOR: TRIACYLGLYCEROL LIPASES

PRELIMINARY AMENDMENT

Assistant Commissioner for Patents  
Washington, DC 20231

Sir:

Before examination of the above-referenced application,  
please amend the application as follows:

IN THE SPECIFICATION:

On page 1, lines 3 and 4, replace the sentence with:

--CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation of International Application No. PCT/US99/09280 filed April 29, 1999, now pending, which claims priority benefit of U.S. Provisional Application No. 60/083,688 filed April 30, 1998.

At page 5, lines 22-23, please delete "an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases.", and insert in its place -- the term "isolated polynucleotide refers to a polynucleotide that is substantially free from other nucleic acid sequences, such as other chromosomal and extrachromosomal DNA and RNA that normally accompany or interact with the isolated polynucleotide as found in its naturally occurring environment.--.

At page 7, line 21, please replace "effecting" with "affecting".

IN THE CLAIMS:

Please cancel claims 1-15 without prejudice or disclaimer.

Please add the following new claims:

--16. An isolated polynucleotide that encodes a polypeptide of at least 80 amino acids, the polypeptide having a sequence identity of at least 80% based on the Clustal method of alignment when compared to a polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34.

17. A polynucleotide sequence of Claim 16, wherein the sequence identity is at least 85%.

18. A polynucleotide sequence of Claim 16, wherein the sequence identity is at least 90%.

19. A polynucleotide sequence of Claim 16, wherein the sequence identity is at least 95%.

20. The polynucleotide of Claim 16 wherein the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34.

21. The polynucleotide of Claim 16, wherein the polynucleotide comprises a nucleotide sequence selected from the group consisting of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, and 33.

22. The polynucleotide of Claim 16, wherein the polypeptide is a triacylglycerol lipase.

23. An isolated complement of the polynucleotide of Claim 16, wherein (a) the complement and the polynucleotide consist of the same number of nucleotides, and (b) the nucleotide sequences of the complement and the polynucleotide have 100% complementarity.

24. An isolated nucleic acid molecule that (1) comprises at least 240 nucleotides and (2) remain hybridized with the isolated polynucleotide of Claim 16 under a wash condition of 0.1X SSC, 0.1% SDS, and 65°C.

25. A cell comprising the polynucleotide of Claim 16.

26. The cell of Claim 25, wherein the cell is selected from the group consisting of a yeast cell, a bacterial cell and a plant cell.

27. A transgenic plant comprising the polynucleotide of Claim 16.

28. A method for transforming a cell comprising introducing into a cell the polynucleotide of Claim 16.

29. A method for producing a transgenic plant comprising (a) transforming a plant cell with the polynucleotide of Claim 16, and (b) regenerating a plant from the transformed plant cell.

30. A method for producing a polynucleotide fragment comprising (a) selecting a nucleotide sequence comprised by the polynucleotide of Claim 16, and (b) synthesizing a polynucleotide fragment containing the nucleotide sequence.

31. The method of Claim 30, wherein the fragment is produced *in vivo*.

32. An isolated polypeptide comprising (a) at least 80 amino acids, and (b) has a sequence identity of at least 80% based on the Clustal method compared to an amino acid sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34.

33. The polypeptide of Claim 32, wherein the sequence identity is at least 85%.

34. The polypeptide of Claim 32, wherein the sequence identity is at least 90%.

35. The polypeptide of Claim 32, wherein the sequence identity is at least 95%.

36. The polypeptide of Claim 32 wherein the polypeptide has a sequence selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32 and 34.

37. The polypeptide of Claim 32, wherein the polypeptide is a triacylglycerol lipase.

38. A chimeric gene comprising the polynucleotide of Claim 16 operably linked to at least one suitable regulatory sequence.

39. A method for altering the level of expression of triacylglycerol lipase in a host cell, the method comprising:

- (a) Transforming a host cell with the chimeric gene of claim 38; and
- (b) Growing the transformed cell in step (a) under conditions suitable for the expression of the chimeric gene.--

## Remarks

Applicants respectfully submit that the amendment to the Specification only clarifies the meaning of the term "isolated" and does not add any new matter. Furthermore, applicants submit that newly added claims more clearly and distinctly recite that which applicants consider to be their invention, and are adequately supported by the original disclosure. For example, the Clustal method of alignment, and the homology percentages are described in the first paragraph on page 5; and the number of 80 amino acids in Claim 1 is supported by the number of nucleotides in SEQ ID NO:16.

No new matter is believed to be at issue. Entry of the amendments and early favorable consideration of the claims on the merits are hereby respectfully requested.

Respectfully submitted,

KENNY LI  
ATTORNEY FOR APPLICANTS  
REGISTRATION NO. 44,872  
TELEPHONE: 302-992-3749  
FACSIMILE: 302-892-1026

Dated: 30 October 2000

**TITLE**  
**TRIACYLGLYCEROL LIPASES**

This application claims the benefit of U.S. Provisional Application No. 60/083,688, filed April 30, 1998.

5

**FIELD OF THE INVENTION**

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding triacylglycerol lipases in plants and seeds.

**BACKGROUND OF THE INVENTION**

10 True lipases attach triacylglycerols and act at an oil-water interface; they constitute a ubiquitous group of enzymes catalyzing a wide variety of reactions, many with industrial potential. Triacylglycerol lipases catalyze the transformation of triacylglycerol and water into diacylglycerol and a fatty acid anion. Human gastric lipase, rat lingual lipase, and human hepatic lysosomal lipase amino acid sequences are homologous but are unrelated to  
15 porcine pancreatic lipase apart from a 6 amino-acid sequence around the essential Ser-152 of porcine pancreatic lipase (Bodmer, M. W. (1987) *Biochim Biophys Acta* 909:237-244). These enzymes are glycosylated, contain a hydrophobic signal peptide, and belong to a gene family of acid lipases (Ameis, D. et al. (1994) *Eur J Biochem* 219:905-914). Lysosomal acid lipase (LAL) is a hydrolase essential for the intracellular degradation of cholesterol esters  
20 and triacylglycerols and participates in the mobilization of seed oil during germination. No plant triacylglycerol lipase cDNAs of this class are currently listed in GenBank.

Neutral triacylglycerol lipases have been widely studied in fungi, bacteria, mammals, and insects. Nucleotide sequences with similarities to neutral triacylglycerol lipases in *Arabidopsis thaliana* and *Ipomea nil* have been described but their function has not yet been proven. The X-ray structure of the *Mucor miehei triglyceride* lipase has been reported, revealing a Ser...His...Asp trypsin-like catalytic triad with an active serine buried under the short helical fragment of a long surface loop (Brady, L. et al. (1990) *Nature* 343:767-770).

It may be useful to isolate triacylglycerol lipase cDNAs from plants that accumulate large amounts of fatty acids with unusual structures. Lacking this ability could be a possible  
30 limitation in development of transgenic crops with novel seed oils. Triacylglycerol lipases may also be useful in processing of plant seed oils. Lysosomal acid lipase (LAL) may be used to engineer lipid and cholesterol ester metabolism and/or lysosome function.

**SUMMARY OF THE INVENTION**

The instant invention relates to isolated nucleic acid fragments encoding  
35 triacylglycerol lipases. Specifically, this invention concerns an isolated nucleic acid fragment encoding an acid or a neutral triacylglycerol lipase. In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding an acid or a neutral triacylglycerol lipase.

An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of a triacylglycerol lipase selected from the group consisting of acid and neutral triacylglycerol lipases.

In another embodiment, the instant invention relates to a chimeric gene encoding an acid or a neutral triacylglycerol lipase, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding an acid or a neutral triacylglycerol lipase, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding an acid or a neutral triacylglycerol lipase, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell.

The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

An additional embodiment of the instant invention concerns a method of altering the level of expression of an acid or a neutral triacylglycerol lipase in a transformed host cell comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding an or a neutral acid triacylglycerol lipase; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of acid or neutral triacylglycerol lipase in the transformed host cell.

An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding an acid or a neutral triacylglycerol lipase.

#### BRIEF DESCRIPTION OF THE

#### DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

Figure 1 depicts the amino acid sequence alignment between the acid triacylglycerol lipase from rice clone rlr72\_pk0015.b2 (SEQ ID NO:14), soybean contig assembled from clones sdp3c\_pk004.n3 and ssl\_pk0022.a1 (SEQ ID NO:18), soybean contig assembled from clones sls1c\_pk009.o2, srl1c\_pk001.m19 and sre\_pk0004.d7 (SEQ ID NO:20), *Canis familiaris* (NCBI General Identifier No. 3041702, SEQ ID NO:35) and *Caenorhabditis elegans* (NCBI General Identifier No. 3165581, SEQ ID NO:36). Amino acids which are conserved among all sequences are indicated with an asterisk (\*) while amino acids

conserved only among plant sequences are indicated by a plus sign (+). Dashes are used by the program to maximize alignment of the sequences.

The following sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent

5 applications as set forth in 37 C.F.R. §1.821-1.825.

SEQ ID NO:1 is the nucleotide sequence comprising the entire cDNA insert in clone cen3n.pk0129.e9 encoding a portion of a corn acid triacylglycerol lipase.

SEQ ID NO:2 is the deduced amino acid sequence of a portion of a corn acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:1.

10 SEQ ID NO:3 is the nucleotide sequence comprising the 3' 647 nucleotides from the cDNA insert in clone ncs.pk0013.h1 encoding the C-terminal quarter of a *Catalpa* acid triacylglycerol lipase

SEQ ID NO:4 is the deduced amino acid sequence of the C-terminal quarter of a *Catalpa* acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:3.

15 SEQ ID NO:5 is the nucleotide sequence comprising the 5' 705 nucleotides from the cDNA insert in clone ncs.pk0013.h1 encoding the N-terminal third of a *Catalpa* acid triacylglycerol lipase.

SEQ ID NO:6 is the deduced amino acid sequence of the N-terminal third of a *Catalpa* acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:5.

20 SEQ ID NO:7 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA insert in clones p0075.cslag33r, p0126.cnlay46r and p0014.ctuty54r encoding a substantial portion of a corn acid triacylglycerol lipase.

SEQ ID NO:8 is the deduced amino acid sequence of a substantial portion of a corn acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:7.

25 SEQ ID NO:9 is the nucleotide sequence comprising a portion of the cDNA insert in clone p0102.ceral64r encoding a portion of a corn acid triacylglycerol lipase.

SEQ ID NO:10 is the deduced amino acid sequence of a portion of a corn acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:9.

30 SEQ ID NO:11 is the nucleotide sequence comprising a portion of the cDNA insert in clone p0126.cn lcm37r encoding a portion of a corn acid triacylglycerol lipase.

SEQ ID NO:12 is the deduced amino acid sequence of a portion of a corn acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:11.

SEQ ID NO:13 is the nucleotide sequence comprising the entire cDNA insert in clone rlr72.pk0015.b2 encoding an entire rice acid triacylglycerol lipase.

35 SEQ ID NO:14 is the deduced amino acid sequence of an entire rice acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:13.

SEQ ID NO:15 is the nucleotide sequence comprising a portion of the cDNA insert in clone rsl1n.pk012.h7 encoding a portion of a rice acid triacylglycerol lipase.

SEQ ID NO:16 is the deduced amino acid sequence of a portion of a rice acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:15.

SEQ ID NO:17 is the nucleotide sequence comprising the contig assembled from the entire cDNA insert in clone ssl.pk0022.a1 and a portion of the cDNA insert in clone 5 sdp3c.pk004.n3 encoding an entire soybean acid triacylglycerol lipase.

SEQ ID NO:18 is the deduced amino acid sequence of an entire soybean acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:17.

SEQ ID NO:19 is the nucleotide sequence comprising the contig assembled from the entire cDNA insert in clone sre.pk0004.d7 and a portion of the cDNA insert in clones 10 sls1c.pk009.o2 and srr1c.pk001.m19 encoding an entire soybean acid triacylglycerol lipase.

SEQ ID NO:20 is the deduced amino acid sequence of an entire soybean acid triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:19.

SEQ ID NO:21 is the nucleotide sequence comprising the entire cDNA insert in clone cr1n.pk0145.c6 encoding half of a corn neutral triacylglycerol lipase.

SEQ ID NO:22 is the deduced amino acid sequence of half of a corn neutral triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:21.

SEQ ID NO:23 is the nucleotide sequence comprising the contig assembled from a portion of the cDNA insert in clones p0010.cbpb40r, p0083.cldeq17r, p0048.cqlac25r, p0118.chsbw59r, cr1.pk0011.c9 and cdo1c.pk002.c22 encoding an entire corn neutral 20 triacylglycerol lipase.

SEQ ID NO:24 is the deduced amino acid sequence of an entire corn neutral triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:23.

SEQ ID NO:25 is the nucleotide sequence comprising the contig assembled from the entire cDNA insert in clone cr1n.pk0127.h8 and a portion of the cDNA insert in clones 25 p0037.crwan02r, p0004.cb1fm22r, p0004.cb1ei43r, cco1n.pk068.o9 and p0093.cssao39r encoding most of a corn neutral triacylglycerol lipase.

SEQ ID NO:26 is the deduced amino acid sequence of most of a corn neutral triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:25.

SEQ ID NO:27 is the nucleotide sequence comprising a portion of the cDNA insert in clone 30 rdr1f.pk002.fl1 encoding a portion of a rice neutral triacylglycerol lipase.

SEQ ID NO:28 is the deduced amino acid sequence of a portion of a rice neutral triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:27.

SEQ ID NO:29 is the nucleotide sequence comprising the contig assembled from the entire cDNA insert in clone sre.pk0058.b1 and a portion of the cDNA insert in clone 35 sah1c.pk001.k20 encoding a substantial portion of a soybean neutral triacylglycerol lipase.

SEQ ID NO:30 is the deduced amino acid sequence of a substantial portion of a soybean neutral triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:29.

SEQ ID NO:31 is the nucleotide sequence comprising the entire cDNA insert in clone sr1.pk0079.e1 encoding the C-terminal half of a soybean neutral triacylglycerol lipase.

SEQ ID NO:32 is the deduced amino acid sequence of the C-terminal half of a soybean neutral triacylglycerol lipase derived from the nucleotide sequence of SEQ ID

5 NO:31.

SEQ ID NO:33 is the nucleotide sequence comprising the entire cDNA insert in clone wr1.pk0115.f5 encoding a portion of a wheat neutral triacylglycerol lipase.

SEQ ID NO:34 is the deduced amino acid sequence of a portion of a wheat neutral triacylglycerol lipase derived from the nucleotide sequence of SEQ ID NO:33.

10 SEQ ID NO:35 is the amino acid sequence of a *Canis familiaris* acid triacylglycerol lipase, NCBI General Identifier No. 3041702.

SEQ ID NO:36 is the amino acid sequence of a *Caenorhabditis elegans* acid triacylglycerol lipase, NCBI General Identifier No. 3165581.

15 The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

20 DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or 25 more segments of cDNA, genomic DNA or synthetic DNA. As used herein, "contig" refers to an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence. For example, several DNA sequences can be compared and aligned to identify common or overlapping regions. The individual sequences can then be assembled into a single contiguous nucleotide sequence.

30 As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration 35 of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional

00000125966960

properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than

5 the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon  
10 encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of  
15 the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with  
20 the nucleic acid fragments disclosed herein.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent similarity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by those skilled in this art. Preferred are those nucleic acid fragments whose nucleotide

25 sequences encode amino acid sequences that are 80% similar to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are 90% similar to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are 95% similar to the amino acid sequences reported herein. Sequence alignments and percent similarity calculations were  
30 performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS*, 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method were  
35 KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and

identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al. (1993) *J. Mol. Biol.* 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide

5 or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as

10 amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a “substantial portion” of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant

15 proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

20 “Codon degeneracy” refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the acid or the neutral triacylglycerol lipase proteins as set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16,

25 18, 20, 22, 24, 26, 28, 30, 32 and 34. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

30 “Synthetic genes” can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. “Chemically synthesized”, as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased

towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding

5 sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived 10 from the same source, but arranged in a manner different than that found in nature.

“Endogenous gene” refers to a native gene in its natural location in the genome of an organism. A “foreign” gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

15 that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a DNA sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation

leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise

30 synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are 35 constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The “translation leader sequence” refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA,

5 mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G. D. (1995) *Molecular Biotechnology* 3:225).

The “3' non-coding sequences” refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The 10 polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al. (1989) *Plant Cell* 1:671-680.

“RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary

15 copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. “Messenger RNA (mRNA)” refers to the RNA that is without introns and that can be translated into protein by the cell. “cDNA” refers to a double-stranded DNA that is complementary to and derived from mRNA. “Sense” RNA 20 refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. “Antisense RNA” refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding 25 sequence, 3' non-coding sequence, introns, or the coding sequence. “Functional RNA” refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term “operably linked” refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For

30 example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term “expression”, as used herein, refers to the transcription and stable

35 accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

“Antisense inhibition” refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. “Overexpression” refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or

non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

5 "Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular 10 localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an

15 amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*)

20 can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of 25 methods of plant transformation include Agrobacterium-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation technology (Klein T. M. et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well 30 known in the art and are described more fully in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several triacylglycerol lipases have been isolated and identified by comparison of random plant cDNA sequences to public 35 databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the proteins that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these proteins.

25996095

TABLE 1  
Triacylglycerol Lipases

Enzyme	Clone	Plant
Triacylglycerol Acid Lipase		
	cen3n.pk0129.e9	Corn
	Contig of: p0075.csrag33r p0126.cnlay46r p0014.ctuty54r	Corn
	p0102.ceral64r	Corn
	p0126.cnlc37r	Corn
	ncs.pk0013.h1	<i>Catalpa</i>
	rlr72.pk0015.b2	Rice
	rsl1n.pk012.h7	Rice
	Contig of: sdp3c.pk004.n3 ssl.pk022.a1	Soybean
	Contig of: s1s1c.pk009.o2 srlc.pk001.m19 sre.pk0004.d7	Soybean
Triacylglycerol Neutral Lipase	cr1n.pk0145.c6	Corn
	Contig of: p0010.cbpbe40r p0083.cldcq17r p0048.cglac25r p0118.chsbw59r cr1.pk0011.c9 cdolc.pk002.c22	Corn
	Contig of: p0037.crwan02r p0004.cb1fm22r p0004.cb1ei43r cc1n.pk068.o9 p0093.cssao39r cr1n.pk0127.h8	Corn
	rdr1f.pk002.fl1	Rice
	Contig of: sah1c.pk001.k20 sre.pk0058.b1	Soybean
	sr1.pk0079.e1	Soybean
	wr1.pk0115.f5	Wheat

5 The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid

hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other acid triacylglycerol lipases, either as cDNAs or

5 genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid  
fragments as DNA hybridization probes to screen libraries from any desired plant employing  
methodology well known to those skilled in the art. Specific oligonucleotide probes based  
upon the instant nucleic acid sequences can be designed and synthesized by methods known  
in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize  
10 DNA probes by methods known to the skilled artisan such as random primer DNA labeling,  
nick translation, or end-labeling techniques, or RNA probes using available *in vitro*  
transcription systems. In addition, specific primers can be designed and used to amplify a  
part or all of the instant sequences. The resulting amplification products can be labeled  
directly during amplification reactions or labeled after amplification reactions, and used as  
15 probes to isolate full length cDNA or genomic fragments under conditions of appropriate  
stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in  
polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding  
homologous genes from DNA or RNA. The polymerase chain reaction may also be

20 performed on a library of cloned nucleic acid fragments wherein the sequence of one primer  
is derived from the instant nucleic acid fragments, and the sequence of the other primer takes  
advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA  
precursor encoding plant genes. Alternatively, the second primer sequence may be based  
upon sequences derived from the cloning vector. For example, the skilled artisan can follow  
25 the RACE protocol (Frohman et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:8998) to generate  
cDNAs by using PCR to amplify copies of the region between a single point in the  
transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed  
from the instant sequences. Using commercially available 3' RACE or 5' RACE systems  
(BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al. (1989) *Proc. Natl.*  
30 *Acad. Sci. USA* 86:5673; Loh et al. (1989) *Science* 243:217). Products generated by the 3'  
and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M. A.  
and Martin, G. R., (1989) *Techniques* 1:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates  
immunological screening of cDNA expression libraries. Synthetic peptides representing

35 portions of the instant amino acid sequences may be synthesized. These peptides can be  
used to immunize animals to produce polyclonal or monoclonal antibodies with specificity  
for peptides or proteins comprising the amino acid sequences. These antibodies can be then  
be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest  
(Lerner, R. A. (1984) *Adv. Immunol.* 36:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed acid or neutral triacylglycerol lipases are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of triacylglycerol and

5 cholesterol esters in those cells. Accumulation of fatty acids with unusual structures may be a positive phenotype in plants used for foods. Triacylglycerol lipases may also be useful in processing of plant seed oils and the development of novel seed oils.

Overexpression of the acid or the neutral triacylglycerol lipases of the instant invention may be accomplished by first constructing a chimeric gene in which the coding

10 region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in 15 order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing 20 the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al.

(1985) *EMBO J.* 4:2411-2418; De Almeida et al. (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of 25 DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant triacylglycerol lipase to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by altering

30 the coding sequence to encode a acid triacylglycerol lipase with appropriate intracellular targeting sequences such as transit sequences (Keegstra, K. (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel, N. (1992) *Plant Phys.* 100:1627-1632) added and/or with targeting sequences that 35 are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding acid or neutral triacylglycerol lipases in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant triacylglycerol lipase can be

constructed by linking a gene or gene fragment encoding an acid or a neutral triacylglycerol lipase to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either

5 the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant acid or neutral triacylglycerol lipases (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be

10 used to prepare antibodies to the these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting acid or neutral triacylglycerol lipases *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant acid or neutral triacylglycerol lipases are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level  
15 expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant acid or neutral triacylglycerol lipase. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded triacylglycerol lipase. An example of a vector for high level expression of the instant acid or  
20 neutral triacylglycerol lipase in a bacterial host is provided (Example 7).

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic  
25 acid fragments may be used as restriction fragment length polymorphism (RFLP) markers.

Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al. (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic  
30 acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein, D. et al. (1980) *Am. J. Hum. Genet.*  
35 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter* 4(1):37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross

populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used  
5 for physical mapping (i.e., placement of sequences on physical maps; *see* Hoheisel, J. D.,  
et al. In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996,  
pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid  
sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask,  
10 B. J. (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor  
use of large clones (several to several hundred KB; *see* Laan, M. et al. (1995) *Genome  
Research* 5:13-20), improvements in sensitivity may allow performance of FISH mapping  
using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical  
15 mapping may be carried out using the instant nucleic acid sequences. Examples include  
allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med.* 114(2):95-96),  
polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics*  
16:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science* 241:1077-1080),  
nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res.* 18:3671), Radiation  
20 Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics* 7:22-28) and Happy Mapping  
(Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods,  
the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in  
the amplification reaction or in primer extension reactions. The design of such primers is  
well known to those skilled in the art. In methods employing PCR-based genetic mapping,  
25 it may be necessary to identify DNA sequence differences between the parents of the  
mapping cross in the region corresponding to the instant nucleic acid sequence. This,  
however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones  
either by targeted gene disruption protocols or by identifying specific mutants for these  
30 genes contained in a maize population carrying mutations in all possible genes (Ballinger  
and Benzer, (1989) *Proc. Natl. Acad. Sci USA* 86:9402; Koes et al. (1995) *Proc. Natl. Acad.  
Sci USA* 92:8149; Bensen et al. (1995) *Plant Cell* 7:75). The latter approach may be  
accomplished in two ways. First, short segments of the instant nucleic acid fragments may  
be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence  
35 primer on DNAs prepared from a population of plants in which Mutator transposons or some  
other mutation-causing DNA element has been introduced (see Bensen, *supra*). The  
amplification of a specific DNA fragment with these primers indicates the insertion of the  
mutation tag element in or near the plant gene encoding the acid or the neutral  
triacylglycerol lipase. Alternatively, the instant nucleic acid fragment may be used as a

hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding an acid or a neutral triacylglycerol lipase can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the acid or the neutral triacylglycerol lipase gene product.

#### EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

#### EXAMPLE 1

##### Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various *Catalpa*, corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

20  
TABLE 2  
cDNA Libraries from Catalpa, Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
ccoln	Corn Cob of 67 Day Old Plants Grown in Green House*	ccoln.pk068.o9
cdolc	Corn Ovary (including pedicel and glumes), 5 Days After Silking	cdolc.pk002.c22
cen3n	Corn Endosperm 20 Days After Pollination*	cen3n.pk0129.e9
cr1	Corn Root From 7 Day Old Seedlings	cr1.pk0011.c9
cr1n	Corn Root From 7 Day Old Seedlings*	cr1n.pk0127.h8 cr1n.pk0145.c6
ncs	<i>Catalpa speciosa</i> Developing Seed	ncs.pk0013.h1
p0004	Corn Immature Ear	p0004.cbliei43r
p0010	Corn Log Phase Suspension Cells Treated With A23187**	p0004.cblfn22r p0010.cbpbe40r
p0014	Corn Leaves 7 and 8 From 3 Foot-Tall Plant	p0014.ctuty54r
p0037	Corn V5 Stage Roots Infested With Corn Root Worm	p0037.crwan02r
p0048	Corn Embryo (Axis and Scutellum) One Day After Germination	p0048.cqlac25r
p0075	Corn Shoot And Leaf Material From Dark-Grown 7 Day-Old Seedlings	p0075.cs slag33r

Library	Tissue	Clone
p0083	Corn Whole Kernels 7 Days After Pollination	p0083.cldcq17r
p0093	Corn Stalk And Shank, 2-3 Weeks After Pollen Shed*	p0093.cssao39r
p0102	Corn Early Meiosis Tassels*	p0102.ceral64r
p0118	Corn Stem Tissue Pooled From the 4 to 5 Internodes Subtending The Tassel At Stages V8-V12, Night Harvested*	p0118.chsbw59r
p0126	Corn Leaf Tissue From V8-V10 Stages, Pooled, Night-Harvested	p0126.cnlay46r p0126.cn lcm37r
rdr1f	Rice Developing Root of 10 Day Old Plant	rdr1f.pk002.f11
rlr72	Rice Leaf 15 Days After Germination, 72 Hours After Infection of Strain <i>Magaporthe grisea</i> 4360-R-62 (AVR2-YAMO); Resistant	rlr72.pk0015.b2
rsl1n	Rice 15-Day-Old Seedling*	rsl1n.pk012.h7
sah1c	Soybean Sprayed With Authority™ Herbicide	sah1c.pk001.k20
sdp3c	Soybean Developing Pods (8-9 mm)	sdp3c.pk004.n3
sls1c	Soybean Infected With <i>Sclerotinia sclerotiorum</i> Mycelium	sls1c.pk009.o2
sr1	Soybean Root	sr1.pk0079.e1
sre	Soybean Root Elongation Zone 4 to 5 Days After Germination	sre.pk0004.d7 sre.pk0058.b1
srr1c	Soybean 8-Day-Old Root	srr1c.pk001.m19
ssl	Soybean Seedling 5-10 Days After Germination	ssl.pk0022.a1
wr1	Wheat Root From 7 Day Old Seedling	wr1.pk0115.f5

\*These libraries were normalized essentially as described in U.S. Patent No. 5,482,845

\*\*A23187 is commercially available from several sources including Calbiochem.

5 cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing

10 recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al. (1991) *Science* 252:1651).

15 The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

ESTs encoding triacylglycerol lipases were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al. (1993) *J. Mol. Biol.* 215:403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)) searches for similarity to sequences contained in the BLAST “nr” database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the “nr” database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the “nr” database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics* 3:266-272) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as “pLog” values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST “hit” represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Acid Triacylglycerol Lipases

The BLASTX search using the EST sequences from clones cen3n.pk0129.e9, ncs.pk0013.h1, a contig sequence assembled from the EST sequences from clones rlr72.pk0015.b2 and rr1.pk0051.f10, a contig sequence assembled from the EST sequences of clones ssl.pk0022.a1 and sr1.pk0098.b11, and a contig sequence assembled from the EST sequences from clones sre.pk0004.d7 and sre.pk0001.b2 revealed similarity of the proteins encoded by the cDNAs and the contigs to acid triacylglycerol lipases from human and rat (GenBank Accession Nos. are listed below). The BLAST results for each of these ESTs and contigs are shown in Table 3:

TABLE 3  
 BLAST Results for Clones Encoding Polypeptides Homologous  
 to Acid Triacylglycerol Lipases

Clone	Organism	GenBank Accession No.	BLAST pLog Score
cen3n.pk0129.e9	Human	X05997	14.52
ncs.pk0013.h1	Rat	X02309	14.70
Contig of rlr72.pk0015.b2 rr1.pk0051.f10	Human	U08464	16.40
Contig of ssl.pk0022.a1 sr1.pk0098.b11	Rat	X02309	15.22
Contig of sre.pk0004.d7 sre.pk0001.b2	Human	X76488	22.00

5        TBLASTN analysis of the proprietary plant EST database indicated that other corn, rice and soybean sequences also encoded acid triacylglycerol lipases. The BLASTX search using the contig sequences assembled with the EST sequences from clones p0075.cslag33r, p0126.cnlay46r and p0014.ctuty54r revealed similarity of the proteins encoded by the cDNAs to acid triacylglycerol lipase from *Homo sapiens* (NCBI General Identifier  
 10 No. 505053). The BLASTX search using the EST sequences from clones p0102.ceral64r and using the contig sequences assembled from the entire cDNA insert in clone  
 ssl.pk0022.a1 and the EST sequences from clone sdp3c.pk004.n3 revealed similarity of the proteins encoded by the cDNAs to acid triacylglycerol lipase from *Canis familiaris* (NCBI General Identifier No. 3041702). The BLASTX search using the EST sequences from clone  
 15 p0126.cnlcml37r revealed similarity of the proteins encoded by the cDNAs to *Drosophila melanogaster* (NCBI General Identifier No. 2894442). The BLASTX search using the EST sequences from clone rsl1n.pk012.h7 revealed similarity of the proteins encoded by the cDNAs to acid triacylglycerol lipase from *Rattus norvegicus* (NCBI General Identifier No. 126307). The BLAST results for each of these sequences is shown in Table 4:  
 20

**TABLE 4**  
 BLAST Results for Clones Encoding Polypeptides Homologous  
 to Acid Triacylglycerol Lipase

Clone	NCBI General Identifier No.	BLAST pLog Score
Contig of: p0075.csflag33r p0126.cnlay46r p0014.ctuty54r	505053	35.00
p0102.ceral64r	3041702	11.30
p0126.cnlem37r	2894442	10.40
rs11n.pk012.h7	126307	7.00

5 The sequence of the entire cDNA insert in clone cen3n.pk0129.e9 was determined and is shown in SEQ ID NO:1; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:2. The amino acid sequence set forth in SEQ ID NO:2 was evaluated by BLASTP, yielding a pLog value of 15.00 versus the *Homo sapiens* sequence (NCBI General Identifier No. 126306). The sequence of the 3'-terminal portion from clone ncs.pk0013.h1 is shown in

10 SEQ ID NO:3; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:4. The sequence of the 5'-terminal portion from clone ncs.pk0013.h1 is shown in SEQ ID NO:5; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:6. The sequence of the contig assembled from the EST sequences from clones p0075.csflag33r, p0126.cnlay46r and p0014.ctuty54r is shown in SEQ ID NO:7; the deduced amino acid

15 sequence of this cDNA is shown in SEQ ID NO:8. The sequence of a portion of the cDNA insert from clone p0102.ceral64r is shown in SEQ ID NO:9; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:10. The sequence of a portion of the cDNA insert from clone p0126.cnlem37r is shown in SEQ ID NO:11; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:12. The sequence of the entire cDNA insert

20 in clone rlr72.pk0015.b2 was determined and is shown in SEQ ID NO:13; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:14. The amino acid sequence set forth in SEQ ID NO:14 was evaluated by BLASTP, yielding a pLog value of 53.30 versus the *C. elegans* sequence (NCBI General Identifier No. 3165581). The sequence of a portion of the cDNA insert from clone rs11n.pk012.h7 is shown in SEQ ID NO:15; the

25 deduced amino acid sequence of this cDNA is shown in SEQ ID NO:16. The sequence of the entire cDNA insert in clone ssi.pk0022.a1 was determined and a contig assembled with this sequence and the EST sequences from clone sdp3c.pk004.n3. The sequence of this contig is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:18. The amino acid sequence set forth in SEQ ID NO:18 was evaluated by

30 BLASTP, yielding a pLog value of 59.40 versus the *C. familiaris* sequence (NCBI General Identifier No. 3041702). The sequence of the entire cDNA insert in clone sre.pk0004.d7 was determined and a contig assembled with this sequence and the EST sequences from

clones *sls1c.pk009.o2* and *srr1c.pk001.m19*. The sequence of this contig is shown in SEQ ID NO:19; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:20. The amino acid sequence set forth in SEQ ID NO:20 was evaluated by BLASTP, yielding a pLog value of 48.70 versus the *C. elegans* sequence (NCBI General Identifier No. 3165581).

5 Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:14, 18 and 20 with the *Canis familiaris* sequence (NCBI General Identifier No. 3041702; SEQ ID NO:35) and the *Caenorhabditis elegans* sequence (NCBI General Identifier No. 3165581; SEQ ID NO:36). The data in Table 5 presents a calculation of the percent similarity of the amino acid sequences set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 10 18 and 20 and the *Caenorhabditis elegans* sequence.

15 **TABLE 5**  
Percent Similarity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Acid Triacylglycerol Lipase

Clone	SEQ ID NO.	3041702	3165581
<i>cen3n.pk0129.e9:fis</i>	2	27.1	22.9
<i>ncs.pk0013.h1:fis1</i>	4	27.4	21.4
<i>ncs.pk0013.h1:fis2</i>	6	30.6	29.9
<i>p0075.cslag33r</i>	8	22.0	23.1
<i>p0126.cnlay46r</i>			
<i>p0014.ctuty54r</i>			
<i>p0102.ceral64r</i>	10	28.8	22.4
<i>p0126.cnlcgm37r</i>	12	26.7	22.2
<i>rlr72.pk0015.b2:fis</i>	14	24.9	25.6
<i>rsl1n.pk012.h7</i>	16	22.5	22.5
<i>sdp3c.pk004.n3</i>	18	27.4	23.1
<i>ssl.pk0022.a1:fis1</i>			
<i>sls1c.pk009.o2</i>	20	23.1	24.8
<i>srr1c.pk001.m19</i>			
<i>sre.pk0004.d7:fis1</i>			

Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the amino acid sequences was performed using the Clustal

20 method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode an entire rice acid triacylglycerol lipase, two different entire soybean acid triacylglycerol lipases, portions from several different corn acid triacylglycerol lipases, portions of a *Catalpa* acid triacylglycerol lipase and a portion of a rice acid

triacylglycerol lipase. These sequences represent the first plant sequences encoding acid triacylglycerol lipases.

EXAMPLE 4

Characterization of cDNA Clones Encoding Neutral Triacylglycerol Lipases

5 The BLASTX search using the contig sequence assembled from the EST sequences from clones cr1n.pk0127.h8 and cr1n.pk0134.d3, and EST sequences from clones cr1n.pk0145.c6, s1.03b01, se3.01a04, sfl1.pk0049.d11, sr1.pk0079.e1, sr1.pk0030.g8, sre.pk0058.b1, wl1n.pk0014.e10, wl1n.pk0038.e3 and wr1.pk0115.f5 revealed similarity of the proteins encoded by the contig and the cDNAs to neutral triacylglycerol lipases from  
10 several organisms. Table 6 shows the BLAST results for the contig and each of the ESTs, the NCBI database accession number, and the organism the closest art sequence is derived from:

15 TABLE 6  
BLAST Results for Clones Encoding Polypeptides Homologous  
to Neutral Triacylglycerol Lipases

Clone	Organism	NCBI Accession No.	BLAST pLog Score
Contig of: cr1n.pk0127.h8 cr1n.pk0134.d3	<i>Thermomyces lanuginosus</i>	999873	10.00
cr1n.pk0145.c6	<i>Caenorhabditis elegans</i>	927399	8.70
sr1.pk0079.e1	<i>Rhizopus niveus</i>	251079	6.70
sre.pk0058.b1	<i>Rhizomucor miehei</i>	417256	8.10
wr1.pk0115.f5	<i>Rhizomucor miehei</i>	82777	6.00

TBLASTN analysis of the proprietary plant EST database indicated that rice clones as well as other corn and soybean clones also encode neutral triacylglycerol lipases. The

20 BLASTX search using the contig sequences assembled from clones p0010.cbpbe40r, p0083.clcdcq17r, p0048.cqlac25r, p0118.chsbw59r, cr1.pk0011.c9 and cdo1c.pk002.c22 and using the EST sequences from clone rdr1f.pk002.f11 revealed similarity of the proteins encoded by the cDNAs to neutral triacylglycerol lipase from *C. elegans* (NCBI General Identifier No.3877256). The BLAST results for each of these sequences are shown in  
25 Table 7:

**TABLE 7**  
 BLAST Results for Clones Encoding Polypeptides Homologous  
 to Neutral Triacylglycerol Lipases

Clone	Organism	NCBI General Identifier No.	BLAST pLog Score
cr1n.pk0145.c6	<i>Caenorhabditis elegans</i>	3877256	9.30
Contig of: p0010.cbpbe40r p0083.cldcq17r p0048.cqlac25r p0118.chsbw59r cr1.pk0011.c9 cdolc.pk002.c22	<i>Caenorhabditis elegans</i>	3877256	18.40
Contig of: p0037.crwan02r p0004.cb1fm22r p0004.cb1ei43r cco1n.pk068.09 p0093.cssao39r cr1n.pk0127.h8	<i>Thermomyces lanuginosus</i>	2997733	6.15
rdr1f.pk002.f11	<i>Caenorhabditis elegans</i>	3877256	10.22
Contig of: sah1c.pk001.k20 sre.pk0058.b1	<i>Rhizomucor miehei</i>	417256	6.22
sr1.pk0079.e1	<i>Rhizopus niveus</i>	3299795	5.70
wr1.pk0115.f5	<i>Caenorhabditis elegans</i>	3877256	14.00

5 The sequence of the entire cDNA insert in clone cr1n.pk0145.c6 was determined and is shown in SEQ ID NO:21; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:22. The amino acid sequence set forth in SEQ ID NO:2 was evaluated by BLASTP, yielding a pLog value of 10.70 versus the *C. elegans* sequence. The sequence of the contig assembled from a portion of the cDNA insert in clones p0010.cbpbe40r,

10 p0083.cldcq17r, p0048.cqlac25r, p0118.chsbw59r, cr1.pk0011.c9 and cdolc.pk002.c22 is shown in SEQ ID NO:23; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:23. The sequence of the entire cDNA insert in clone cr1n.pk0127.h8 was determined and a contig assembled with this sequence and the sequence from a portion of the cDNA insert in clones p0037.crwan02r, p0004.cb1fm22r, p0004.cb1ei43r, cco1n.pk068.09 and p0093.cssao39r. The sequence of this contig is shown in SEQ ID NO:25; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:26. The amino acid sequence set forth in SEQ ID NO:4 was evaluated by BLASTP, yielding a pLog value of 9.70 versus the *Thermomyces lanuginosus* sequence. The sequence of a portion of the cDNA insert from clone rdr1f.pk002.f11 is shown in SEQ ID NO:27; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:28. The sequence of the entire cDNA insert in clone sre.pk0058.b1 was determined and a contig assembled with this sequence and the sequence

of a portion of the cDNA insert in clone sah1c.pk001.k20. The sequence of this contig is shown in SEQ ID NO:29; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:30. The amino acid sequence set forth in SEQ ID NO:30 was evaluated by BLASTP, yielding a pLog value of 8.05 versus the *Rhizomucor miehei* sequence. The

5 sequence of the entire cDNA insert in clone sr1.pk0079.e1 was determined and is shown in SEQ ID NO:31; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:32. The amino acid sequence set forth in SEQ ID NO:32 was evaluated by BLASTP, yielding a pLog value of 7.52 versus the *Rhizopus niveus* sequence. The sequence of the entire cDNA insert in clone wr1.pk0115.f5 was determined and is shown in SEQ ID NO:33; the deduced  
10 amino acid sequence of this cDNA is shown in SEQ ID NO:34. The amino acid sequence set forth in SEQ ID NO:34 was evaluated by BLASTP, yielding a pLog value of 13.52 versus the *Caenorhabditis elegans* sequence.

The data in Table 8 presents a calculation of the percent similarity of the amino acid sequences set forth in SEQ ID NOs:22, 24, 26, 28, 30, 32 and 34 and the *Caenorhabditis elegans*, *Rhizomucor miehei* and *Thermomyces lanuginosus* sequences.

TABLE 8

Percent Similarity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Neutral Triacylglycerol Lipase

Clone	SEQ ID NO.	3877256	2997733	417256
cr1n.pk0145.c6	22	15.1	13.2	16.8
Contig of:	24	60.5	17.5	22.9
p0010.cbpbe40r				
p0083.cldcq17r				
p0048.cqlac25r				
p0118.chsbw59r				
cr1.pk0011.c9				
cdolc.pk002.c22				
Contig of:	26	18.5	14.3	15.1
p0037.crwan02r				
p0004.cb1fm22r				
p0004.cb1ei43r				
ccoln.pk068.o9				
p0093.cssao39r				
cr1n.pk0127.h8				
rdr1f.pk002.f11	28	12.6	20.6	22.9
Contig of:	32	15.1	10.5	17.0
sah1c.pk001.k20				
sre.pk0058.b1				
sr1.pk0079.e1	34	14.3	21.1	24.6
wr1.pk0115.f5	36	37.0	22.0	26.0

Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the amino acid sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS*. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10).

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode three different corn neutral triacylglycerol lipases (one portion and two entire or nearly entire), two different soybean triacylglycerol lipases (one portion and one nearly entire) and a portion of a wheat triacylglycerol lipase. These sequences represent the first monocot and soybean sequences encoding neutral triacylglycerol lipases.

#### EXAMPLE 5

##### Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding triacylglycerol lipases in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA

fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (Nco I or Sma I) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below.

Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes Nco I and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb Nco I-Sma I

fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sal I-Nco I promoter fragment of the maize 27 kD zein gene and a 0.96 kb Sma I-Sal I fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli*

XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment

encoding a triacylglycerol lipase, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed

with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al. (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

5 The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers 10 resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

15 The particle bombardment method (Klein, T. M. et al. (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1  $\mu$ m in diameter) are coated with DNA using the following technique. Ten  $\mu$ g of plasmid DNAs are added to 50  $\mu$ L of a suspension of gold particles (60 mg per mL). Calcium chloride (50  $\mu$ L of a 2.5 M solution) and spermidine free base (20  $\mu$ L of a 1.0 M solution) 20 are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200  $\mu$ L of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30  $\mu$ L of ethanol. An aliquot (5  $\mu$ L) of the DNA-coated 25 gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of 30 about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock 35 tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains glufosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter

of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

5 Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al. (1990) *Bio/Technology* 8:833-839).

#### EXAMPLE 6

##### Expression of Chimeric Genes in Dicot Cells

10 A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the  $\beta$  subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant triacylglycerol lipase in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon 15 and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

20 The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

25 Soybean embryos may then be transformed with the expression vector comprising sequences encoding a triacylglycerol lipase. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and 30 placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule.

35 Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Klein T. M. et al. (1987) *Nature* (London) 327:70-73, U.S.

Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al.

5 (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the triacylglycerol lipase and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then

10 be inserted into a unique restriction site of the vector carrying the marker gene.

To 50  $\mu$ L of a 60 mg/mL 1  $\mu$ m gold particle suspension is added (in order): 5  $\mu$ L DNA (1  $\mu$ g/ $\mu$ L), 20  $\mu$ L spermidine (0.1 M), and 50  $\mu$ L CaCl<sub>2</sub> (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400  $\mu$ L 70%

15 ethanol and resuspended in 40  $\mu$ L of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five  $\mu$ L of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette.

20 For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

25 Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into

30 individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

#### EXAMPLE 7

##### Expression of Chimeric Genes in Microbial Cells

The cDNAs encoding the instant triacylglycerol lipases can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and

Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using 5 oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve 10

GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase 15 (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent 20 cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the triacylglycerol lipase are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) 25 (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM 30 DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One µg of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating 35 at the expected molecular weight.

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding all or a substantial portion of an acid triacylglycerol lipase comprising a member selected from the group consisting of:
  - 5 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18 and SEQ ID NO:20;
  - 10 (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18 and SEQ ID NO:20; and
  - 15 (c) an isolated nucleic acid fragment that is complementary to (a) or (b).
2. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17 and SEQ ID NO:19.
3. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.
4. A transformed host cell comprising the chimeric gene of Claim 3.
5. An acid triacylglycerol lipase polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18 and SEQ ID NO:20.
6. An isolated nucleic acid fragment encoding all or a substantial portion of a neutral triacylglycerol lipase comprising a member selected from the group consisting of:
  - 30 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 and SEQ ID NO:34;
  - (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 and SEQ ID NO:34; and

(c) an isolated nucleic acid fragment that is complementary to (a) or (b).

7. The isolated nucleic acid fragment of Claim 6 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:21, SEQ ID NO:23, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:31 and SEQ ID NO:33.

5 8. A chimeric gene comprising the nucleic acid fragment of Claim 6 operably linked to suitable regulatory sequences.

9. A transformed host cell comprising the chimeric gene of Claim 8.

10 10. A neutral triacylglycerol lipase polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:22, SEQ ID NO:24, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:32 and SEQ ID NO:34.

15 11. A method of altering the level of expression of a triacylglycerol lipase in a host cell comprising:

15 (a) transforming a host cell with the chimeric gene of any of Claims 3 and 8; and

(b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene

20 wherein expression of the chimeric gene results in production of altered levels of a triacylglycerol lipase in the transformed host cell.

20 12. A method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a triacylglycerol lipase comprising:

25 (a) probing a cDNA or genomic library with the nucleic acid fragment of any of Claims 1 and 6;

(b) identifying a DNA clone that hybridizes with the nucleic acid fragment of any of Claims 1 and 6;

(c) isolating the DNA clone identified in step (b); and

(d) sequencing the cDNA or genomic fragment that comprises the clone isolated in step (c)

30 30 wherein the sequenced nucleic acid fragment encodes all or a substantial portion of the amino acid sequence encoding a triacylglycerol lipase.

35 13. A method of obtaining a nucleic acid fragment encoding a substantial portion of an amino acid sequence encoding a triacylglycerol lipase comprising:

35 (a) synthesizing an oligonucleotide primer corresponding to a portion of the sequence set forth in any of SEQ ID NOs: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31 and 33; and

(b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a) and a primer representing sequences of the cloning vector

wherein the amplified nucleic acid fragment encodes a substantial portion of an amino acid sequence encoding a triacylglycerol lipase.

14. The product of the method of Claim 12.
15. The product of the method of Claim 13.

TITLE

TRIACYLGLYCEROL LIPASES

ABSTRACT OF THE DISCLOSURE

This invention relates to an isolated nucleic acid fragment encoding a triacylglycerol 5 lipase. The invention also relates to the construction of a chimeric gene encoding all or a portion of the triacylglycerol lipase, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the triacylglycerol lipase in a transformed host cell.

10

15

20

25

30

35

KL/dmm

EXPRESS MAIL LABEL NO. EK639605137US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

#### IN THE APPLICATION OF:

E. I. DUPONT DE NEMOURS AND COMPANY

CASE NO.: BB1168 US NA

APPLICATION NO.: UNKNOWN

GROUP ART UNIT: UNKNOWN

FILED: CONCURRENTLY HEREWITH

EXAMINER: UNKNOWN

FOR: TRIACYLGLYCEROL LIPASES

Assistant Commissioner for Patents  
Washington, DC 20231

Sir:

**DECLARATION IN ACCORDANCE WITH 37 CFR 1.821**

I hereby state that the content of the paper and computer readable copies of the Sequence Listing, submitted in accordance with 37 CFR 1.821(c) and (e), respectively are the same.

Respectfully submitted,

KENING LI  
ATTORNEY FOR APPLICANTS  
REGISTRATION NO. 44,872  
TELEPHONE: 302-992-3749  
FACSIMILE: 302-892-1026

Dated: 30 October 2000

SEQUENCE LISTING

Val Tyr Arg Met Ser Gly Ile Pro Pro Ser Phe Pro Leu Phe Leu Ser  
 65 70 75 80

Tyr Gly Gly Arg Asp Ser Leu Ala Asp Pro Ala Asp Val Arg Leu Leu  
85 90 95

Leu Gln Asp Leu Arg Gly His Asp Gln Asp Lys Leu Thr Val Gln Tyr  
100 105 110

Leu Asp Lys Phe Ala His Leu Asp Phe Ile Ile Gly Val Cys Ala Lys  
 115 120 125

Asp Tyr Val Tyr Lys Asp Met Ile Asp Phe Leu Asn Arg Phe Asn  
 130 135 140

<210> 3  
<211> 647  
<212> DNA  
<213> *Catalpa* sp.

<210> 4  
<211> 116  
<212> PRT  
<213> *Catalpa* sp.

<400> 4  
Ile Glu Thr Pro Val His Met Ala Gln Thr Val Arg Tyr Gly Ile Leu  
1 5 10 15

Pro Lys Tyr Asp Tyr Gly Asn Pro Ser Phe Asn Leu Ala His Tyr Gly  
20 25 30

Glu Ser Arg Pro Pro Val Tyr Asp Leu Ser Lys Ile Pro Leu Asp Ile  
35 40 45

Pro Leu Phe Leu Ser Tyr Gly Gly Gln Asp Ala Leu Ser Asp Val Lys  
50 55 60

Asp Val Glu Thr Leu Leu Asp Ser Leu Lys Leu His Asp Val Asp Lys  
 65 70 75 80

Leu His Val Gln Tyr Ile Lys Asp Tyr Ala His Ala Asp Phe Ile Ile  
 85 90 95

Gly Val Thr Ala Lys Asp Ile Val Tyr Asn Gln Ile Val Thr Phe Phe  
100 105 110

Arg Asn Gln Ala  
115

<210> 5  
<211> 705  
<212> DNA  
<213> Catalpa sp.

<220>  
<221> unsure  
<222> (526)

<220>  
<221> unsure  
<222> (561)

<220>  
<221> unsure  
<222> (585)

<220>  
<221> unsure  
<222> (593)

<220>  
<221> unsure  
<222> (664)

<220>  
<221> unsure  
<222> (679)

<400> 5  
gcacgagcca acagcttctt aaatttagct ctcttaatcc ttctctcatt atcaactactc 60  
cttccatc aatcattcgc ctccagccgc cgccgttttc ttccgcagaa tgatgtcggtt 120  
cttccgcgg acggcgtttt ctccacccgc gtaactgtac acggttataa atgccaagaa 180  
tttgcgatgtt caactgtatgtatggatataa ttaagctgtgc agaggattttt ggagggccgg 240  
ggccggaggag gaggggcgaa gggccgcgg gttctgtgtc aacatgggggt ttctgtggac 300  
gggtatgacgt ggctggtaa tggacccgaa caatctttgg cgtatgtatattt ggctgacaaat 360  
gggttcgcgg tctggatttc taacataaga ggaacttaggtt ttagtctgtgc tcatgtcaac 420  
cttgcattca cggatcgttca atattggat tgggcattggg acgtatctgg tgacccacaga 480  
cttacatcc ctgtatcgatgtatggat ttagtggatcg acaaaacgggtt cagaanacac actacatatagg 540  
gcaatccatc gggaaactttt nttagctttgg gatcactttt aggangggaaa cangttggca 600  
gggttaatcg gctgtatgtt aagccaaattt gctaaacgggtt catatgcacac tgctctcgag 660  
ttgncatggaa gatccttggat gggaaacaca cgtatcttggc ctgcgtt 705

<210> 6  
<211> 157  
<212> PRT  
<213> Catalpa sp.

<400> 6  
Ala Arg Ala Asn Ser Phe Leu Asn Leu Ala Leu Leu Ile Leu Leu Ser  
1 5 10 15

Leu Ser Leu Leu Leu Pro His Gln Ser Phe Ala Ser Ser Arg Arg Arg  
20 25 30

Phe Leu Pro Gln Asn Asp Val Val Leu Pro Pro Asp Gly Val Cys Ser  
35 40 45

Thr Ala Val Thr Val His Gly Tyr Lys Cys Gln Glu Phe Glu Val Thr  
 50 55 60

Thr Asp Asp Gly Tyr Ile Leu Ser Val Gln Arg Ile Leu Glu Gly Arg  
65 70 75 80

Ala Gly Gly Gly Pro Lys Arg Pro Pro Val Leu Leu Gln His Gly  
 85 90 95

Val Leu Val Asp Gly Met Thr Trp Leu Val Asn Gly Pro Glu Gln Ser  
100 105 110

Leu Ala Met Ile Leu Ala Asp Asn Gly Phe Asp Val Trp Ile Ser Asn  
115 120 125

Ile Arg Gly Thr Arg Phe Ser Arg Arg His Val Ser Leu Asp Pro Thr  
130 135 140

Asp Pro Glu Tyr Trp Asp Trp Ala Trp Asp Asp Leu Gly  
145 150 155

<210> 7  
<211> 859  
<212> DNA  
<213> Zea mays

<220>  
<221> unsure  
<222> (46)

<220>  
<221> unsure  
<222> (231)

```

<400> 7
aaacccaaaac acggcgccaca tggtgccgccc aggaaaaagcg ctgcgcgcgc cccagcttc 60
cttcctcggt ttccctgcgc tcctagccgg tggagccgcg gcatacccgcc ccacagacgc 120
gtcaacggccg gtctcccccgg ggcggggggc cgggtggcctc tgccacgcgc tgctctcgcc 180
gcagggttacc cgtgcacccga gcacacccatc acaacggatgc atggtttttc ntgttcctt 240
cagcatattc cacaatggcgc aataatggaaatt gcacataataa tggacacttc agttttttttt 300
cagccacggtc ttttccgggg tggagataca tggttccatata actccaaatgc aacaatccatc 360
ggatataatcc ttgtgcacaaatc ttgtttttgtat tggtttttgcgaaatgttcg tggcacacgt 420
ttggatataag gccactatcac tctctctgtt catgataaagc ttttctggaa tttggatgttgc 480
caacggatctt ctgaatagca cttttttggca atgtttaagct atgtatatac atgttccatc 540
tccaaatattt ttgtatgtggg acatccacag gaaactatca tgggttttgc tgctgttaca 600
atgcctgtaaa cagttaaatgt gataaaatctt gttggcgttcc ttgtttccat tttcttaccc 660
gatacgttgc gtgtctgtt ttgttttaga gcaatgttgc tgcatacttgc tgatgttgc tgatgttgc 720
gttattatgg gcatccatca ttgttgcatttcc cggagcgatata tgggttttgcata gatattatgtat 780
tcgtctgtggc atgtatgttgc ttgttgcacttgc aacatgttgc tatcttcaat aacatgttgc 840
acttttttttcc aatcatctc 859

```

<210> 8  
<211> 286  
<212> PRT  
<213> *Zea mays*

<220>  
<221> UNSURE  
<222> (16)

<400> 8  
Lys Ala Asn Asn Gly Gly His Gly Ala Pro Arg Lys Ser Ala Cys Xaa  
1 5 10 15

Ala Pro Ala Pro Pro Pro Arg Val Pro Leu Pro Pro Ser Arg Trp Ser  
20 25 30

Pro Arg Ile Pro Ala His Arg Arg Ala Thr Pro Arg Leu Pro Ala Arg  
35 40 45

Gly Gly Arg Trp Pro Leu Pro Ala Ala Ala Pro Ala Ala Gly Tyr Pro  
50 55 60

Cys Thr Glu His Thr Val Gln Thr Asp Asp Gly Phe Leu Leu Ser Leu  
65 70 75 80

Gln His Ile Pro His Gly Arg Asn Gly Ile Ala Asp Asn Thr Gly Pro  
85 90 95

Pro Val Phe Leu Gln His Gly Leu Phe Gln Gly Gly Asp Thr Trp Phe  
100 105 110

Ile Asn Ser Asn Glu Gln Ser Leu Gly Tyr Ile Leu Ala Asp Asn Gly  
115 120 125

Phe Asp Val Trp Val Gly Asn Val Arg Gly Thr Arg Trp Ser Lys Gly  
130 135 140

His Ser Thr Leu Ser Val His Asp Lys Leu Phe Trp Asp Trp Ser Trp  
145 150 155 160

Gln Asp Leu Ala Glu Tyr Asp Val Leu Ala Met Leu Ser Tyr Val Tyr  
165 170 175

Thr Val Ala Gln Ser Lys Ile Leu Tyr Val Gly His Ser Gln Gly Thr  
180 185 190

Ile Met Gly Leu Ala Ala Phe Thr Met Pro Glu Thr Val Lys Met Ile  
195 200 205

Ser Ser Ala Ala Leu Leu Cys Pro Ile Ser Tyr Leu Asp His Val Ser  
210 215 220

Ala Ser Phe Val Leu Arg Ala Val Ala Met His Leu Asp Glu Met Leu  
225 230 235 240

Val Ile Met Gly Ile His Gln Leu Asn Phe Arg Ser Asp Met Gly Val  
245 250 255

Gln Ile Leu Asp Ser Leu Cys Asp Asp Glu His Leu Asp Cys Asn Asp  
260 265 270

Leu Leu Ser Ser Ile Thr Val Lys Thr Val Val Gln Ser Ser  
275 280 285

<210> 9  
<211> 509  
<212> DNA  
<213> Zea mays

卷之三

<220>  
<221> unsure  
<222> (162)

<220>  
<221> unsure  
<222> (277)

<220>  
<221> unsure  
<222> (284)

<220>  
<221> unsure  
<222> (290)

<220>  
<221> unsure  
<222> (295)

<220>  
<221> unsure  
<222> (386)

<220>  
<221> unsure  
<222> (406)

<220>  
<221> unsure  
<222> (413)

<220>  
<221> unsure  
<222> (443)

<220>  
<221> unsure  
<222> (468)

<220>  
<221> unsure  
<222> (484)

<220>  
<221> unsure  
<222> (489)

<400> 9  
cgatcgagat ggctcagaag gatctctate taccgttctt ggcttcttcc atcattgcct 60  
gctgcttgat gaacctgaa agtgttctca gtcagaagcag gatgcggaaat actacaaacg 120  
atattatggta tgacaaatgc ccccccacaac ctcatccctt angtatgtgc aggtcccgag 180  
tagcagctta cggctatcctt tgtagggaaat accatgtgac agggggat ggtctacatcc 240  
ttatcttaaa gaagatcccc tatgttcttctt ctggtnacac cganattaaacn agganacccg 300  
tactactgtt ccatggctta ctgggtggat gtttctgttgg gtgtactgaac acacaaaaac 360  
aatactctggg cttcacccctgg ctgaatggggat tttggatggat ggtatcggatc tcnccccggaa 420  
aaaatccacc gagggacaca ccnctccccaaa aaaaacccggc tttgggangg aatggaaacac 480  
tgcnaaaaaa actcccgctgt gtcgaatcc 509

<210> 10  
<211> 125

<212> PRT  
<213> Zea mays

<220>  
<221> UNSURE  
<222> (52)

<220>  
<221> UNSURE  
<222> (90)

<220>  
<221> UNSURE  
<222> (92)

<220>  
<221> UNSURE  
<222> (96)

<400> 10  
Met Ala Gln Lys Asp Leu Tyr Leu Pro Phe Leu Ala Leu Ser Ile Ile  
1 5 10 15

Ala Cys Cys Leu Met Asn Leu Gln Ser Val Leu Ser Ser Ser Arg Met  
20 25 30

Arg Asn Thr Thr Asn Asp Ile Ser Asp Asp Lys Cys Pro Pro Gln Pro  
                  35                 40                 45

His Pro Leu Xaa Met Cys Arg Ser Arg Val Ala Ala Tyr Gly Tyr Pro  
50 55 60

Cys Glu Glu Tyr His Val Thr Thr Glu Asp Gly Tyr Ile Leu Ser Leu  
 65 70 75 80

Lys Lys Ile Pro Tyr Gly Leu Ser Gly Xaa Thr Xaa Ile Thr Arg Xaa  
85 90 95

Pro Val Leu Leu Phe His Gly Leu Leu Val Asp Gly Phe Cys Trp Val  
100 105 110

115 120 125

<211> 273  
<212> DNA  
<213> Zea mays

<221> unsure  
<222> (8)

<221> unsure  
<222> (20)

<221> unsure  
<222> (229)

<220>  
 <221> unsure  
 <222> (236)

<220>  
 <221> unsure  
 <222> (241)

<220>  
 <221> unsure  
 <222> (249)

<220>  
 <221> unsure  
 <222> (268)

<400> 11  
 ctcttcntg cacgcttgn tttcagctct actggaaactg gtcctggat gacctggtag 60  
 tcaacgacatg gccggccatg gtcgacttgc tcgtcaaca gacccggccag aacccctcaact 120  
 acgtcgaca ctccatggg acgtcggtgg cgctggcgc ctctcgag ggccgggtgg 180  
 tgagccatg gaaatcccg gcgctgtcga cgccgtgtgc ctacctcgnca cacatnaaca 240  
 nccccaaatng gaatcttgtt tggccaaangc gtt 273

<210> 12  
 <211> 90  
 <212> PRT  
 <213> Zea mays

<220>  
 <221> UNSURE  
 <222> (76)

<220>  
 <221> UNSURE  
 <222> (78)

<220>  
 <221> UNSURE  
 <222> (80)

<220>  
 <221> UNSURE  
 <222> (83)

<220>  
 <221> UNSURE  
 <222> (89)

<400> 12  
 Ser Ser Cys Thr Leu Arg Phe Gln Leu Tyr Trp Asn Trp Ser Trp Asp  
 1 5 10 15

Asp Leu Val Val Asn Asp Leu Pro Ala Met Val Asp Phe Val Val Lys  
 20 25 30

Gln Thr Gly Gln Lys Pro His Tyr Val Gly His Ser Met Gly Thr Leu  
 35 40 45

Val Ala Leu Ala Ala Phe Ser Glu Gly Arg Val Val Ser Gin Leu Lys  
 50 55 60

Ser Ala Ala Leu Leu Thr Pro Val Ala Tyr Leu Xaa His Xaa Asn Xaa  
 65 70 75 80

Pro Asn Xaa Asn Pro Gly Trp Pro Xaa Arg  
85 90

<210> 13

<211> 1483

<212> DNA

<213> *Oryza sativa*

<400> 13

52103 14

**210** 14

211 410  
212 RBT

<212> PRI  
<213> *Oryza sativa*

<400> 14

Met Ala Met Ala Gly His Ala Pro Gly Gly Ala Leu Pro Leu Ile Leu  
1 5 10 15

Leu Val Val Ser Cys Cys Gly Arg Ile Val Ser Gly Ala Ser Pro Ala  
20 25 30

Ala Ala Ala Leu Arg Arg Val Gly Ser Gly Ser Gly Gly Leu Cys Asp  
35 40 45

Gln Leu Leu Leu Pro Leu Gly Tyr Pro Cys Thr Glu His Asn Val Glu  
50 55 60

Thr Lys Asp Gly Phe Leu Leu Ser Leu Gln His Ile Pro His Gly Lys  
65 70 75 80

Asn Lys Ala Ala Asp Ser Thr Gly Pro Pro Val Phe Leu Gln His Gly  
                  85                 90                 95

Leu Phe Gln Gly Gly Asp Thr Trp Phe Ile Asn Ser Ala Glu Gln Ser  
 100 105 110  
 Leu Gly Tyr Ile Leu Ala Asp Asn Gly Phe Asp Val Trp Ile Gly Asn  
 115 120 125  
 Val Arg Gly Thr Arg Trp Ser Lys Gly His Ser Thr Phe Ser Val His  
 130 135 140  
 Asp Lys Leu Phe Trp Asp Trp Ser Trp Gln Glu Leu Ala Glu Tyr Asp  
 145 150 155 160  
 Leu Leu Ala Met Leu Gly Tyr Val Thr Val Thr Gln Ser Lys Ile  
 165 170 175  
 Leu Tyr Val Gly His Ser Gln Gly Thr Ile Met Gly Leu Ala Ala Leu  
 180 185 190  
 Thr Met Pro Glu Ile Val Lys Met Ile Ser Ser Ala Ala Leu Leu Cys  
 195 200 205  
 Pro Ile Ser Tyr Leu Asp His Val Ser Ala Ser Phe Val Leu Arg Ala  
 210 215 220  
 Val Ala Met His Leu Asp Gln Met Leu Val Thr Met Gly Ile His Gln  
 225 230 235 240  
 Leu Asn Phe Arg Ser Asp Met Gly Val Gln Ile Val Asp Ser Leu Cys  
 245 250 255  
 Asp Gly Glu His Val Asp Cys Asn Asn Leu Leu Ser Ala Ile Thr Gly  
 260 265 270  
 Glu Asn Cys Cys Phe Asn Thr Ser Arg Ile Asp Tyr Tyr Leu Glu Tyr  
 275 280 285  
 Glu Pro His Pro Ser Ser Thr Lys Asn Leu His His Leu Phe Gln Met  
 290 295 300  
 Ile Arg Lys Gly Thr Phe Ala Lys Tyr Asp Tyr Gly Leu Leu Gly Asn  
 305 310 315 320  
 Leu Arg Arg Tyr Gly His Leu Arg Pro Pro Ala Phe Asp Leu Ser Ser  
 325 330 335  
 Ile Pro Glu Ser Leu Pro Ile Trp Met Gly Tyr Gly Gly Leu Asp Ala  
 340 345 350  
 Leu Ala Asp Val Thr Asp Val Gln Arg Thr Ile Arg Glu Leu Gly Ser  
 355 360 365  
 Thr Pro Glu Leu Leu Tyr Ile Gly Asp Tyr Gly His Ile Asp Phe Val  
 370 375 380  
 Met Ser Val Lys Ala Lys Asp Asp Val Tyr Val Asp Leu Ile Arg Phe  
 385 390 395 400  
 Leu Arg Glu Asn Gly Trp His Asn Ser Tyr  
 405 410

<210> 15  
<211> 395  
<212> DNA  
<213> *Oryza sativa*

<220>  
<221> unsure  
<222> (12)

<220>  
<221> unsure  
<222> (24)

<220>  
<221> unsure  
<222> (29)

<220>  
<221> unsure  
<222> (33)

<220>  
<221> unsure  
<222> (43)

<220>  
<221> unsure  
<222> (78)

<220>  
<221> unsure  
<222> (182)

<220>  
<221> unsure  
<222> (265)

<220>  
<221> unsure  
<222> (300)

<220>  
<221> unsure  
<222> (302)

<220>  
<221> unsure  
<222> (306)

<220>  
<221> unsure  
<222> (347)

<220>  
<221> unsure  
<222> (351)

<220>  
<221> unsure  
<222> (367)

```

<220>
<221> unsure
<222> (370)

<220>
<221> unsure
<222> (380)

<220>
<221> unsure
<222> (386)

<400> 15
acatcttca cnggaaaaa ctantgtcng aanaattcag canccgacat cttcctcaag 60
tacgagcccc agccaaatnc cacaaaaaac ttgatccatc tcgctcaaac ggtgagagac 120
gggttctga ccaagatcga ctacgtatc ccggacgcga acgtggccag gtacggcag 180
gncgaccgcg cggctacga catggcgcg atcccgccgt gttcccccatttccctcagc 240
tacgccggcc gggactcgct gtccnacccc gccgatcgct gcctctcc tcgacgatcn 300
cngccnccg ggcacatcg ggcacccgtc catccgtcc agtaacnttc nccatactcg 360
cccacgnccn acttcgtcan tcgggnnttc tgccg 395

<210> 16
<211> 80
<212> PRT
<213> Oryza sativa

<220>
<221> UNSURE
<222> (8)

<220>
<221> UNSURE
<222> (10)..(11)

<220>
<221> UNSURE
<222> (15)

<220>
<221> UNSURE
<222> (61)

<400> 16
Thr Ser Phe Thr Gly Lys Asn Xaa Cys Xaa Xaa Asn Ser Ala Xaa Asp
1 5 10 15
Ile Phe Leu Lys Tyr Glu Pro Gln Pro Thr Ser Thr Lys Thr Leu Ile
20 25 30
His Leu Ala Gln Thr Val Arg Asp Gly Val Leu Thr Lys Tyr Asp Tyr
35 40 45
Val Met Pro Asp Ala Asn Val Ala Arg Tyr Gly Gln Xaa Asp Pro Pro
50 55 60
Ala Tyr Asp Met Ala Ala Ile Pro Ala Trp Phe Pro Ile Phe Leu Ser
65 70 75 80

<210> 17
<211> 1718

```

<212> DNA  
<213> Glycine max

<400> 17

<210> 18

<210> 40

<211> 410  
<212> PRT  
<213> Glycine max

<400> 18

Met. Al. L.

Val Leu Thr Thr Val Pro Arg Gln Ala His Ala Ser Ser Ser Arg Gly Asn  
20 25 30

35 40 45

50 55 60

65                    70                    75                    80  
 Ser Ser Gly Ser Gly Thr Arg Lys Gln Pro Val Val Ile Gln His Gly

Val Leu Val Asp Gly Met Thr Trp Leu Leu Asn Pro Pro Glu Gln Asp

Leu Pro Leu Ile Leu Ala Asp Asn Gly Phe Asp Val Trp Ile Ala Asn  
 115 120 125  
 Thr Arg Gly Thr Arg Tyr Ser Arg Arg His Ile Ser Leu Asp Pro Ser  
 130 135 140  
 Ser Gln Ala Tyr Trp Asn Trp Ser Trp Asp Glu Leu Val Ser Tyr Asp  
 145 150 155 160  
 Phe Pro Ala Val Phe Asn Tyr Val Phe Ser Gln Thr Gly Gln Lys Ile  
 165 170 175  
 Asn Tyr Val Gly His Ser Leu Gly Thr Leu Val Ala Leu Ala Ser Phe  
 180 185 190  
 Ser Glu Gly Lys Leu Val Thr Gln Leu Lys Ser Ala Ala Leu Leu Ser  
 195 200 205  
 Pro Ile Ala Tyr Leu Ser His Met Asn Thr Ala Leu Gly Val Val Ala  
 210 215 220  
 Pro Lys Ser Phe Val Gly Glu Ile Thr Thr Leu Phe Gly Leu Ala Glu  
 225 230 235 240  
 Phe Asn Pro Lys Gly Leu Ala Val Asp Ala Phe Leu Lys Ser Leu Cys  
 245 250 255  
 Ala His Pro Gly Ile Asp Cys Tyr Asp Leu Leu Thr Ala Leu Thr Gly  
 260 265 270  
 Lys Asn Cys Cys Leu Asn Ser Ser Thr Val Asp Leu Phe Leu Met Asn  
 275 280 285  
 Glu Pro Gln Ser Thr Ser Thr Lys Asn Met Val His Leu Ala Gln Thr  
 290 295 300  
 Val Arg Leu Gly Ala Leu Thr Lys Phe Asn Tyr Val Arg Pro Asp Tyr  
 305 310 315 320  
 Asn Ile Met His Tyr Gly Glu Ile Phe Pro Pro Ile Tyr Asn Leu Ser  
 325 330 335  
 Asn Ile Pro His Asp Leu Pro Leu Phe Ile Ser Tyr Gly Arg Asp  
 340 345 350  
 Ala Leu Ser Asp Val Arg Asp Val Glu Asn Leu Leu Asp Lys Leu Lys  
 355 360 365  
 Phe His Asp Glu Asn Lys Arg Ser Val Gln Phe Ile Gln Glu Tyr Ala  
 370 375 380  
 His Ala Asp Tyr Ile Met Gly Phe Asn Ala Lys Asp Leu Val Tyr Asn  
 385 390 395 400  
 Ala Val Leu Ser Phe Phe Asn His Gln Val  
 405 410  
 <210> 19  
 <211> 1438  
 <212> DNA  
 <213> Glycine max

<210> 20  
<211> 405  
<212> PRT  
<213> Glycine max

<400> 20  
Met Arg Ile Gln Arg Phe Leu Ala Thr Leu Ala Ile Thr Val Ser Ile

Leu Leu Gly Asn Gly Asn Pro Val Gln Cys Phe Asp Gly Gly Ser His

Gln Lys Gln Gln His Ser Leu Cys Glu Glu Leu Ile Ile Pro Tyr Gly  
25 30 35 40 45

Tyr Pro Cys Ser Glu His Thr Ile Gln Thr Lys Asp Gly Phe Leu Leu  
50 55 60

Gly Leu Gln Arg Val Ser Ser Ser Ser Ser Leu Arg Leu Arg Asn His  
65 70 75 80

Gly Asp Gly Gly Pro Pro Val Leu Leu Leu His Gly Leu Phe Met Ala  
85 90 95

Gly Asp Ala Trp Phe Leu Asn Thr Pro Glu Gln Ser Leu Gly Phe Ile  
100 105 110

Leu Ala Asp His Gly Phe Asp Val Trp Val Gly Asn Val Arg Gly Thr  
115 120 125

Arg Trp Ser His Gly His Ile Ser Leu Leu Glu Lys Lys Lys Gln Phe  
130 135 140

Trp Asp Trp Ser Trp Gln Glu Leu Ala Leu Tyr Asp Val Ala Glu Met  
 145 150 155 160  
 Ile Asn Tyr Ile Asn Ser Val Thr Asn Ser Lys Ile Phe Val Val Gly  
 165 170 175  
 His Ser Gln Gly Thr Ile Ile Ser Leu Ala Ala Phe Thr Gln Pro Glu  
 180 185 190  
 Ile Val Glu Lys Val Glu Ala Ala Leu Leu Ser Pro Ile Ser Tyr  
 195 200 205  
 Leu Asp His Val Ser Ala Pro Leu Val Leu Arg Met Val Lys Met His  
 210 215 220  
 Ile Asp Glu Met Ile Leu Thr Met Gly Ile His Gln Leu Asn Phe Lys  
 225 230 235 240  
 Ser Glu Trp Gly Ala Ser Leu Leu Val Ser Leu Cys Asp Thr Arg Leu  
 245 250 255  
 Ser Cys Asn Asp Met Leu Ser Ser Ile Thr Gly Lys Asn Cys Cys Phe  
 260 265 270  
 Asn Glu Ser Arg Val Glu Phe Tyr Leu Glu Gln Glu Pro His Pro Ser  
 275 280 285  
 Ser Ser Lys Asn Leu Asn His Leu Phe Gln Met Ile Arg Lys Gly Thr  
 290 295 300  
 Tyr Ser Lys Tyr Asp Tyr Gly Lys Leu Lys Asn Leu Ile Glu Tyr Gly  
 305 310 315 320  
 Lys Phe Asn Pro Pro Lys Phe Asp Leu Ser Arg Ile Pro Lys Ser Leu  
 325 330 335  
 Pro Leu Trp Met Ala Tyr Gly Asn Asp Ala Leu Ala Asp Ile Thr  
 340 345 350  
 Asp Phe Gln His Thr Leu Lys Glu Leu Pro Ser Pro Pro Glu Val Val  
 355 360 365  
 Tyr Leu Glu Asn Tyr Gly His Val Asp Phe Ile Leu Ser Leu Gln Ala  
 370 375 380  
 Lys Gln Asp Leu Tyr Asp Pro Met Ile Ser Phe Phe Lys Ser Ser Gly  
 385 390 395 400  
 Lys Phe Ser Ser Met  
 405

<210> 21  
 <211> 737  
 <212> DNA  
 <213> Zea mays

<400> 21  
 gcacgaggtt ttgtgccctt gatctatctg ttaaattttgg atcgcaggag gttgaactca 60  
 tgaccttgg acagcctcgg ataggcaatc ctgcatttgc tttatctttt ggtgaacaag 120  
 tcccaagaac aatccgtgtg acccatcaga atgatattgt gccgcattta ccaccgtatt 180  
 attttacatc aggtgaatgg acataccacc acttcgcgtatc agagggtttgg cttcatgaga 240

```

gcatagatgg  aaatgttagt  accagaaacg  agacggatg  tgatgattct  ggtaaagacc  300
cgccatcttag  cagggtcgtc  tatggatgg  ggcgttgcaga  tcatctttgt  tactatctgt  360
tcacactat  tcgtgttttca  agaggaaacct  gtcatttcgt  gatgttgcc  gcaaccaaa  420
tatatacaacta  cgttcgtgaa  gttgtatggat  cccatcatctt  gtcaagatac  cccgaagaa  480
caacaaggctt  aagaatctatgt  tgactttgtt  tgccacggaa  tgccatgtttt  tacatgttt  540
tttccatttca  tttttgttgc  agtcattgtt  atgttttttttt  ctctttggcc  ttccatggcc  600
ataaggagag  gtcacccctt  ttaaatgttc  cccctttgtt  caagtggaaaa  tcgtgtcatgt  660
aagctccatca  agatgttccg  cacaatttca  ttgtgttata  taaaataatac  tatgtgttac  720
aaaaaaaaaa  aaaaaaaaaa  aaaaaaaaaa  aaaaaaaaaa  aaaaaaaaaa  aaaaaaaaaa  737

```

<210> 22  
<211> 166  
<212> PRT  
<213> Zea mays

<400> 22

Thr Arg Phe Cys Ala Leu Asp Leu Ser Val Lys Phe Gly Ser Gln Glu  
 1 5 10 15

Val Glu Leu Met Thr Phe Gly Gln Pro Arg Ile Gly Asn Pro Ala Phe  
20 25 30

Ala Val Tyr Phe Gly Glu Gln Val Pro Arg Thr Ile Arg Val Thr His  
35 40 45

Gln Asn Asp Ile Val Pro His Leu Pro Pro Tyr Tyr Tyr Tyr Tyr Leu Gly  
50 55 60

Glu Trp Thr Tyr His His Phe Ala Arg Glu Val Trp Leu His Glu Ser  
 65 70 75 80

Ile Asp Gly Asn Val Val Thr Arg Asn Glu Thr Val Cys Asp Asp Ser  
85 90 95

Gly Glu Asp Pro Thr Cys Ser Arg Ser Val Tyr Gly Met Ser Val Ala  
100 105 110

Asp His Leu Glu Tyr Tyr Asp Val Thr Leu His Ala Asp Ser Arg Gly  
115 120 125

Thr Cys Gin Phe Val Ile Gly Ala Ala Asn Gin Val Tyr Asn Tyr Val  
130 135 140

Arg Glu Val Asp Gly Ser Ile Ile Leu Ser Arg Tyr Pro Gin Glu Pro  
145 150 155 160

Gin Ala Leu Glu Ser Met

<210> 23  
<211> 1434  
<212> DNA  
<213> Zea mays

<220>  
<221> unsure  
<222> (226)

<220>  
<221> unsure



<220>  
<221> UNSURE  
<222> (129)

<400> 24  
Met Glu Arg Trp Ser Leu Gly Ala Lys Val Val Ala Leu Ala Leu Leu  
1 5 10 15  
  
Leu Ser Ala Ala Ser His Gly Arg Glu Leu Pro Val Lys Ser Ser Asp  
20 25 30  
  
Arg Ser Phe Ile Tyr Asn His Thr Leu Ala Lys Thr Leu Val Glu Tyr  
35 40 45  
  
Ala Xaa Ala Val Tyr Met Thr Asp Leu Thr Ala Leu Phe Thr Trp Thr  
50 55 60  
  
Cys Ser Arg Cys Asn Asp Leu Thr Gln Gly Phe Glu Met Arg Ser Xaa  
65 70 75 80  
  
Ile Val Asp Val Glu Lys Leu Leu Ala Gly Ile Val Gly Val Asp His  
85 90 95  
  
Ser Leu Asn Ser Ile Ile Val Ala Ile Arg Gly Thr Gln Glu Asn Ser  
100 105 110  
  
Val Gln Asn Trp Ile Lys Asp Leu Ile Trp Lys Gln Leu Asp Leu Ser  
115 120 125  
  
Xaa Pro Asn Met Pro Asn Ala Lys Val His Ser Gly Phe Phe Ser Ser  
130 135 140  
  
Tyr Asn Asn Thr Ile Leu Arg Leu Ala Ile Thr Ser Ala Val His Lys  
145 150 155 160  
  
Ala Arg Lys Ser Tyr Gly Asp Ile Asn Val Ile Val Thr Gly His Ser  
165 170 175  
  
Met Gly Gly Ala Met Ala Ser Phe Cys Ala Leu Asp Leu Ala Met Lys  
180 185 190  
  
Leu Gly Gly Ser Val Gln Leu Met Thr Phe Gly Gln Pro Arg Val  
195 200 205  
  
Gly Asn Ala Ala Phe Ala Ser Tyr Phe Ala Lys Tyr Val Pro Asn Thr  
210 215 220  
  
Ile Arg Val Thr His Gly His Asp Ile Val Pro His Leu Pro Pro Tyr  
225 230 235 240  
  
Phe Ser Phe Leu Pro Gln Leu Thr Tyr His His Phe Pro Arg Glu Val  
245 250 255  
  
Trp Val Gln Asp Ser Asp Gly Asn Thr Thr Glu Arg Ile Cys Asp Asp  
260 265 270  
  
Ser Gly Glu Asp Pro Asp Cys Cys Arg Cys Ile Ser Met Phe Gly Leu  
275 280 285  
  
Arg Ile Gln Asp His Ser Leu Thr  
290 295



Oryza sativa L.

Val Gln Lys Gln Leu Gln Lys His Pro Asn Ala Asn Val Val Val Thr  
 65 70 75 80

Gly His Ser Leu Gly Ala Ala Leu Ala Thr Ile Phe Pro Ala Leu Leu  
 85 90 95

Ala Phe His Gly Glu Arg Gly Val Leu Asp Arg Leu Leu Ser Val Val  
 100 105 110

Thr Tyr Gly Gln Pro Arg Val Gly Asp Lys Val Phe Ala Gly Tyr Val  
 115 120 125

Arg Ala Asn Val Pro Val Glu Pro Leu Arg Val Val Tyr Arg Tyr Asp  
 130 135 140

Val Val Pro Arg Val Pro Phe Asp Ala Pro Pro Val Ala Asp Phe Ala  
 145 150 155 160

His Gly Gly Thr Cys Val Tyr Phe Asp Gly Trp Tyr Lys Gly Arg Glu  
 165 170 175

Ile Ala Lys Gly Gly Asp Ala Pro Asn Lys Asn Tyr Phe Asp Pro Arg  
 180 185 190

Tyr Leu Leu Ser Met Tyr Gly Asn Ala Trp Gly Asp Leu Phe Lys Gly  
 195 200 205

Ala Phe Leu Trp Ala Lys Glu Gly Lys Asp Tyr Arg Glu Gly Ala Val  
 210 215 220

Ser Leu Leu Tyr Arg Ala Thr Gly Leu Leu Val Pro Gly Ile Ala Ser  
 225 230 235 240

His Ser Pro Arg Asp Tyr Val Asn Ala Val Arg Leu Gly Ser Val Ala  
 245 250 255

Ser Ala

<210> 27  
 <211> 432  
 <212> DNA  
 <213> Oryza sativa

<220>  
 <221> unsure  
 <222> (7)

<220>  
 <221> unsure  
 <222> (15)

<220>  
 <221> unsure  
 <222> (27)

<220>  
 <221> unsure  
 <222> (38)

<220>  
 <221> unsure  
 <222> (50)

<220>  
 <221> unsure  
 <222> (94)

<220>  
 <221> unsure  
 <222> (99)

<220>  
 <221> unsure  
 <222> (103)

<220>  
 <221> unsure  
 <222> (105)

<220>  
 <221> unsure  
 <222> (117)

<400> 27  
 catagtnata atacaacag ttgcggncat tgagatntt ggaatctgn tcggggca 60  
 aggaagacat atggaggct acctataaat gttnttagnt canthcgatg ggagggncct 120  
 tttagcatcg ttctttgtcc cttagccctt ctgttaaagt atggatcgca ggaagtcaa 180  
 ctcactgact ttggacagcc tcgggttaggc aatccctt ttgctgcgta cttcagtgac 240  
 caagtccgca gaacaatccg tttgaccatc cagaatgaca ttgtcccaactt cttggccacca 300  
 tatttttgttgc accttggcga atggacatcat caccacttct cttggagggat ttggcttcat 360  
 gagccatcat taggaaatgt agttacttagg aatggaccca tctgtatgg atcaggcgag 420  
 gaccaacat gc 432

<210> 28  
 <211> 106  
 <212> PRT  
 <213> Oryza sativa

<400> 28  
 Gly Pro Phe Ser Ile Val Leu Val Pro Leu Thr Ser Leu Val Lys Tyr  
 1 5 10 15

Gly Ser Gln Glu Val Gln Leu Met Thr Phe Gly Gln Pro Arg Val Gly  
 20 25 30

Asn Pro Ser Phe Ala Ala Tyr Phe Ser Asp Gln Val Pro Arg Thr Ile  
 35 40 45

Arg Val Thr His Gln Asn Asp Ile Val Pro His Leu Pro Pro Tyr Phe  
 50 55 60

Cys Tyr Leu Gly Glu Trp Thr Tyr His His Phe Ser Arg Glu Val Trp  
 65 70 75 80

Leu His Glu Thr Ile Val Gly Asn Val Val Thr Arg Asn Glu Thr Ile  
 85 90 95

Cys Asp Gly Ser Gly Glu Asp Pro Thr Cys  
 100 105

<210> 29  
<211> 1234  
<212> DNA  
<213> Glycine max

<400> 29  
ccactggaaatggaatttcg tgagatttt tgattgtcgg gaatgattt caagaaaagg 60  
ccacaaccca agtcttgatt gttttggaca agcatgagaa ccgcgatact tatgtggtag 120  
cttcccgagg aacggaaaccc ttgtatgcag atgcatgtgc cactgaccc gacatctcat 180  
ggtagccatt cccggcattt gaaaaatgca tggtggcttc atgaaagccct tagggctaca 240  
gaaaaatgt gggttggctta aggagatcca aagggtgaa aatcttcccc cttggcccta 300  
ctatgttat agggacattc taaggaaagg ttgtgtgaa aatcctaattg caaagtttat 360  
cattacgggt catatgtttgg gtggagactc cgaatcttgc taccctacga tcatgttct 420  
gcatgtatgg aagtggctga ttgagagggtt ggaaggatc tacacgtttg ggcaaccaag 480  
agttggagat gaagcatatc cacatgtatg gagacaaaaaa ttgagggaaa attctatcag 540  
gtatggcagg ttgtttttat gcaatgcattt agtccggagg ttgcccattt atgataaagg 600  
cttgcattt aagactttt ggatctgcct ttcttttaac aggcgcattt aactcaggat 660  
tctcgaagaa gaggcgaata agaaacttattt ctgcgcattt tggtgtatc ccatgtatgtt 720  
caatgtgtt tggactta taaggagctt taccatagcg tacaaaaaatg gacctacta 780  
tagagaagga tggttttctt ttatgttcag ttgtgtgtt ctgtgttcc ctggcttacc 840  
tgcgtacccgtt ccacaaatg atataatc cactcttcg ggtcaatcggaaaatcattt 900  
taaagcagat tgatgttcc gtatacatga tcattccata ccactacgtt catgtgtatg 960  
gtcatgcaga cttaaatata cataatcaag attttttatgg ttagaaaaaaa tggtaataaa 1020  
actttgttattt gtatcatgtt aagaatagtgtt atgtatcata atgtatcatgataatataac 1080  
agtttgcgtt cagtagcgtt tattttatgg taatataaa gcttaggttta aagttgtttc 1140  
ctttgtgtca tggatttttattt ataaatgttcaatgttca gtttgggttattt ttcaaaaaaaa 1200  
aaaaaaaaaaa aaaaaaaaaaaa aaaaaaaa 1234

<210> 30  
<211> 246  
<212> PRT  
<213> Glycine max

<400> 30  
His Leu Met Val Arg Ile Pro Gly Ile Gly Lys Met His Gly Gly Phe  
1 5 10 15  
Met Lys Ala Leu Gly Leu Gln Lys Asn Val Gly Trp Pro Lys Glu Ile  
20 25 30  
Gln Arg Asp Glu Asn Leu Pro Pro Leu Ala Tyr Tyr Val Ile Arg Asp  
35 40 45  
Ile Leu Arg Lys Gly Leu Ser Glu Asn Pro Asn Ala Lys Phe Ile Ile  
50 55 60  
Thr Gly His Ser Leu Gly Gly Ala Leu Ala Ile Leu Tyr Pro Thr Ile  
65 70 75 80  
Met Phe Leu His Asp Glu Lys Leu Leu Ile Glu Arg Leu Glu Gly Ile  
85 90 95  
Tyr Thr Phe Gly Gln Pro Arg Val Gly Asp Glu Ala Tyr Ala Gln Tyr  
100 105 110  
Met Arg Gln Lys Leu Arg Glu Asn Ser Ile Arg Tyr Cys Arg Phe Val  
115 120 125  
Tyr Cys Asn Asp Ile Val Pro Arg Leu Pro Tyr Asp Asp Lys Asp Leu  
130 135 140

Leu Phe Lys His Phe Gly Ile Cys Leu Phe Asn Arg Arg Tyr Glu  
145 150 155 160

Leu Arg Ile Leu Glu Glu Glu Pro Asn Lys Asn Tyr Phe Ser Pro Trp  
165 170 175

Cys Val Ile Pro Met Met Phe Asn Ala Val Leu Glu Leu Ile Arg Ser  
180 185 190

Phe Thr Ile Ala Tyr Lys Asn Gly Pro His Tyr Arg Glu Gly Trp Phe  
195 200 205

Leu Phe Ser Phe Arg Leu Val Gly Leu Leu Ile Pro Gly Leu Pro Ala  
210 215 220

His Gly Pro Gln Asp Tyr Ile Asn Ser Thr Leu Leu Gly Ser Ile Glu  
225 230 235 240

Lys His Phe Lys Ala Asp  
245

<210> 31

<211> 490

<212> DNA

<213> Glycine max

<400> 31

gcacgaggag agatggccta aagaatttga aaccgtatgg aaccgtccac gtgtcttatta 60  
ctccataagg gatgttctaa aagaatgtt gataatggaaat gataaaagccaa agtttattttt 120  
tacgggtatc agtcttggtg gagcacttgc aattctttt cccgctatgc taattttgc 180  
tgctgagaca tttcttttgg aaaggcttgg aagggtgtac acatttggac agccctagggt 240  
tggatggatc aacattttgtca aatacatggaa aatcatatgg aacacattatgg gatattttttt 300  
ttttaggtt gtttactgtca acgtatattttt tccttaggtt cccttttgatc aagatatacat 360  
gaaattttggat cattttggga catgtcttta ttatgacagg agctatacat gcaaggtaaca 420  
tataataatgtt ttaatttttt ttgattcatg catatatttcg tcattgttaat caactttttt 480  
ttttctgggg 490

<210> 32

<211> 141

<212> PRT

<213> Glycine max

<400> 32

His Glu Glu Arg Trp Pro Lys Glu Ile Glu Thr Asp Glu Asn Arg Pro  
1 5 10 15

Arg Val Tyr Tyr Ser Ile Arg Asp Leu Leu Lys Lys Cys Leu Asn Arg  
20 25 30

Asn Asp Lys Ala Lys Phe Ile Leu Thr Gly His Ser Leu Gly Gly Ala  
35 40 45

Leu Ala Ile Leu Phe Pro Ala Met Leu Ile Leu His Ala Glu Thr Phe  
50 55 60

Leu Leu Glu Arg Leu Glu Gly Val Tyr Thr Phe Gly Gln Pro Arg Val  
65 70 75 80

Gly Asp Glu Thr Phe Ala Lys Tyr Met Glu Asn Gln Leu Lys His Tyr  
85 90 95

9  
8  
7  
6  
5  
4  
3  
2  
1  
0

Gly Ile Lys Tyr Phe Arg Phe Val Tyr Cys Asn Asp Ile Val Pro Arg  
100 105 110

Leu Pro Phe Asp Glu Asp Ile Met Lys Phe Glu His Phe Gly Thr Cys  
115 120 125

Leu Tyr Tyr Asp Arg Ser Tyr Thr Cys Lys Val His Ile  
130 135 140

<210> 33  
<211> 774  
<212> DNA  
<213> Triticum aestivum

<400> 33  
gcacgagaat attccatca tggcacagg acattccatg ggaggggca tggcttcgtt 60  
ttgtgcctt gatcttattc tcaactatgg gttaaaggac gtgaccctgc tgacattttg 120  
gcaaccccggtt atggatgtg tccggatcc taccacttta aagaataact tgccaaacgc 180  
aattcgatgtt accaaccgcac atgatattgt gcctcatca ccccccgtactt accagttttt 240  
cccacacaaat accttccatc atttccacc acagggtttgg gttcataaca ttggactcga 300  
tagctacta taccggatcg agcacatcg tgatcatttgg gggaaaagac cccacttgc 360  
ggaggccctt ggttggaaat agcgtccagg cccatcaccc ctttcttgc tcaggatccc 420  
atcccgatgtc ggcggatca tccggatcg tcacggatcg caatatgc tggcacaacg 480  
ttggccctt agacgggtgtt attgtctt cgaagcggcc tggttatca gtttgcgttgc 540  
tactcgtatc acatggaaataa agctcaagat tacatggatt tattttgttgc 600  
ccaaagaaaca atttctgtt tggcaatccaa agcactatctt catgtatata taccgcgtgtc 660  
atccctggctt gattaaatata tcctagctga ggggttattt ctgaaatgtt ccaacatata 720  
tatgtgttgcatt aaaaaaaaaaaaaaaatatac ttgaggcgcc cccgttccaaat 774

<210> 34  
<211> 126  
<212> PRT  
<213> Triticum aestivum

<400> 34  
His Glu Asn Ile Pro Ile Met Val Thr Gly His Ser Met Gly Gly Ala  
1 5 10 15

Met Ala Ser Phe Cys Ala Leu Asp Leu Ile Val Asn Tyr Gly Leu Lys  
20 25 30

Asp Val Thr Leu Leu Thr Phe Gly Gln Pro Arg Ile Gly Asn Ala Val  
35 40 45

Phe Ala Thr His Phe Lys Tyr Leu Pro Asn Ala Ile Arg Val Thr  
50 55 60

Asn Ala His Asp Ile Val Pro His Leu Pro Pro Tyr Tyr Gln Tyr Phe  
65 70 75 80

Pro Gln Asn Thr Tyr His His Phe Pro Pro Glu Val Trp Val His Asn  
85 90 95

Ile Gly Leu Asp Ser Leu Leu Tyr Pro Ile Glu His Ile Cys Asp His  
100 105 110

Ser Gly Glu Arg Pro His Leu Gln Gln Ala Leu Gly Trp Lys  
115 120 125

<210> 35  
<211> 398

<212> PRT  
<213> Canis familiaris

<400> 35  
Met Trp Leu Leu Leu Thr Ala Ala Ser Val Ile Ser Thr Leu Gly Thr  
1 5 10 15  
Thr His Gly Leu Phe Gly Lys Leu His Pro Thr Asn Pro Glu Val Thr  
20 25 30  
Met Asn Ile Ser Gln Met Ile Thr Tyr Trp Gly Tyr Pro Ala Glu Glu  
35 40 45  
Tyr Glu Val Val Thr Glu Asp Gly Tyr Ile Leu Gly Ile Asp Arg Ile  
50 55 60  
Pro Tyr Gly Arg Lys Asn Ser Glu Asn Ile Gly Arg Arg Pro Val Ala  
65 70 75 80  
Phe Leu Gln His Gly Leu Leu Ala Ser Ala Thr Asn Trp Ile Ser Asn  
85 90 95  
Leu Pro Asn Asn Ser Leu Ala Phe Ile Leu Ala Asp Ala Gly Tyr Asp  
100 105 110  
Val Trp Leu Gly Asn Ser Arg Gly Asn Thr Trp Ala Arg Arg Asn Leu  
115 120 125  
Tyr Tyr Ser Pro Asp Ser Val Glu Phe Trp Ala Phe Ser Phe Asp Glu  
130 135 140  
Met Ala Lys Tyr Asp Leu Pro Ala Thr Ile Asp Phe Ile Leu Lys Lys  
145 150 155 160  
Thr Gly Gln Asp Lys Leu His Tyr Val Gly His Ser Gln Gly Thr Thr  
165 170 175  
Ile Gly Phe Ile Ala Phe Ser Thr Asn Pro Lys Leu Ala Lys Arg Ile  
180 185 190  
Lys Thr Phe Tyr Ala Leu Ala Pro Val Ala Thr Val Lys Tyr Thr Glu  
195 200 205  
Thr Leu Leu Asn Lys Leu Met Leu Val Pro Ser Phe Leu Phe Lys Leu  
210 215 220  
Ile Phe Gly Asn Lys Ile Phe Tyr Pro His His Phe Phe Asp Gln Phe  
225 230 235 240  
Leu Ala Thr Glu Val Cys Ser Arg Glu Thr Val Asp Leu Leu Cys Ser  
245 250 255  
Asn Ala Leu Phe Ile Ile Cys Gly Phe Asp Thr Met Asn Leu Asn Met  
260 265 270  
Ser Arg Leu Asp Val Tyr Leu Ser His Asn Pro Ala Gly Thr Ser Val  
275 280 285  
Gln Asn Val Leu His Trp Ser Gln Ala Val Lys Ser Gly Lys Phe Gln  
290 295 300

Ala Phe Asp Trp Gly Ser Pro Val Gln Asn Met Met His Tyr His Gln  
 305 310 315 320  
 Ser Met Pro Pro Tyr Tyr Asn Leu Thr Asp Met His Val Pro Ile Ala  
 325 330 335  
 Val Trp Asn Gly Gly Asn Asp Leu Leu Ala Asp Pro His Asp Val Asp  
 340 345 350  
 Leu Leu Leu Ser Lys Leu Pro Asn Leu Ile Tyr His Arg Lys Ile Pro  
 355 360 365  
 Pro Tyr Asn His Leu Asp Phe Ile Trp Ala Met Asp Ala Pro Gln Ala  
 370 375 380  
 Val Tyr Asn Glu Ile Val Ser Met Met Gly Thr Asp Asn Lys  
 385 390 395  
 <210> 36  
 <211> 403  
 <212> PRT  
 <213> Caenorhabditis elegans  
 <400> 36  
 Met Trp Arg Phe Ala Val Phe Leu Ala Ala Phe Phe Val Gln Asp Val  
 1 5 10 15  
 Val Gly Ser His Gly Asp Pro Glu Leu His Met Thr Thr Pro Gln Ile  
 20 25 30  
 Ile Glu Arg Trp Gly Tyr Pro Ala Met Ile Tyr Thr Val Ala Thr Asp  
 35 40 45  
 Asp Gly Tyr Ile Leu Glu Met His Arg Ile Pro Phe Gly Lys Thr Asn  
 50 55 60  
 Val Thr Trp Pro Asn Gly Lys Arg Pro Val Val Phe Met Gln His Gly  
 65 70 75 80  
 Leu Leu Cys Ala Ser Ser Asp Trp Val Val Asn Leu Pro Asp Gln Ser  
 85 90 95  
 Ala Gly Phe Leu Phe Ala Asp Ala Gly Phe Asp Val Trp Leu Gly Asn  
 100 105 110  
 Met Arg Gly Asn Thr Tyr Ser Met Lys His Lys Asp Leu Lys Pro Ser  
 115 120 125  
 His Ser Ala Phe Trp Asp Trp Ser Trp Asp Glu Met Ala Thr Tyr Asp  
 130 135 140  
 Leu Asn Ala Met Ile Asn His Val Leu Glu Val Thr Gly Gln Asp Ser  
 145 150 155 160  
 Val Tyr Tyr Met Gly His Ser Gln Gly Thr Leu Thr Met Phe Ser His  
 165 170 175  
 Leu Ser Lys Asp Asp Gly Ser Phe Ala Lys Lys Ile Lys Lys Phe Phe  
 180 185 190

Ala Leu Ala Pro Ile Gly Ser Val Lys His Ile Lys Gly Phe Leu Ser  
 195 200 205  
 Phe Phe Ala Asn Tyr Phe Ser Leu Glu Phe Asp Gly Trp Phe Asp Ile  
 210 215 220  
 Phe Gly Ala Gly Glu Phe Leu Pro Asn Asn Trp Ala Met Lys Leu Ala  
 225 230 235 240  
 Ala Lys Asp Ile Cys Gly Leu Lys Val Glu Ala Asp Leu Cys Asp  
 245 250 255  
 Asn Val Leu Phe Leu Ile Ala Gly Pro Glu Ser Asp Gln Trp Asn Gln  
 260 265 270  
 Thr Arg Val Pro Val Tyr Ala Thr His Asp Pro Ala Gly Thr Ser Thr  
 275 280 285  
 Gln Asn Ile Val His Trp Met Gln Met Val His His Gly Gly Val Pro  
 290 295 300  
 Ala Tyr Asp Trp Gly Thr Lys Thr Asn Lys Lys Lys Tyr Gly Gin Ala  
 305 310 315 320  
 Asn Pro Pro Glu Tyr Asp Phe Thr Ala Ile Lys Gly Thr Lys Ile Tyr  
 325 330 335  
 Leu Tyr Trp Ser Asp Ala Asp Trp Leu Ala Asp Thr Pro Asp Val Pro  
 340 345 350  
 Asp Tyr Leu Leu Thr Arg Leu Asn Pro Ala Ile Val Ala Gln Asn Asn  
 355 360 365  
 His Leu Pro Asp Tyr Asn His Leu Asp Phe Thr Trp Gly Leu Arg Ala  
 370 375 380  
 Pro Asp Asp Ile Tyr Arg Pro Ala Ile Lys Leu Cys Thr Asp Asp Tyr  
 385 390 395 400  
 Leu Gly Lys